

Chapter 1

INTRODUCTION AND OVERVIEW

[1.1]	M. V. Berry, “The Quantum Phase, Five Years After”*	7
[1.2]	R. Jackiw, “Three Elaborations on Berry’s Connection, Curvature and Phase,” <i>Int. J. Mod. Phys. A3</i> (1988) 285–297	29

* Original Contribution.

This page is intentionally left blank

1

Introduction and Overview

The two articles contained in this chapter form the proper introduction to the main contents of the book. Here, we briefly discuss the general notion of holonomy, and illustrate it with a few primeval examples.

A phase is, for our purposes, not a state of matter but a complex number of unit modulus, an element of the group $U(1)$. We shall use the term somewhat loosely to encompass elements of matrix groups as well, such as $U(N)$. The phases we shall be interested in are often associated with cyclic evolution of a physical system. More specifically, we shall find that the cyclic variation of external parameters often leads to a net evolution involving a phase depending only on the *geometry* of the path traversed in parameter space. In other words, this phase is independent of how fast the various parts of the path are traversed. For non-cyclic evolution, the extra phase will depend on the endpoints of the path. The phase is non-integrable; it can not be written as a function just of the endpoints, because it depends on the geometry of the path connecting them as well.

The natural mathematical context for geometric phases is the theory of $U(N)$ fiber bundles. There one defines a phase, known as a holonomy,* that depends on the geometry of a loop, and is independent of any coordinate choice. (For a brief introduction to the world of fiber bundles, connections, and holonomy, see the introduction to Chapter 3.)

Examples of geometric phases abound in many areas of physics. Many familiar problems that we do not ordinarily associate with geometric phases may be phrased in terms of them. Often, the result is a clearer understanding of the structure of the problem, and an elegant expression of its solution.

Consider, for example, the precession of a Foucault pendulum. Standard treatments¹ calculate the rate of precession of a pendulum in a frame rotating with the surface of the earth in terms of the Coriolis force, but a much simpler and more geometric explanation may be given as follows. Suppose

* This phase is also, and perhaps more properly, known as an *anholonomy*. We shall adhere to the established mathematical terminology, although this other usage is quite common in the physics literature.

a pendulum is transported along a closed loop C , in the gravitational field of a point mass, and that the period and amplitude of its swing are small compared to the typical time and distance scales of the transport motion. We may assume that the loop lies on the surface of a sphere concentric with the mass, although this assumption is not necessary. Now when the pendulum returns to its initial position, its invariant plane will have rotated by some angle. For example, for transport around the sphere at a constant latitude of θ (relative to the north pole), a straightforward calculation shows that the net rotation will be $2\pi \cos \theta$ radians. A remarkable feature of this result is that it is independent of the rate at which different parts the loop are traversed (provided that the traversal is slow). This is a consequence of the fact that the Coriolis force is proportional to the velocity of transport, so that its integrated effect is invariant under rescalings of time. (Velocity-dependent forces, like the magnetic force on a moving charge, tend to be associated with geometric phases.) How does the pendulum precess as it is taken around a general path C ? For transport along the equator, the pendulum will not precess. This may be seen from a symmetry argument. The rate of precession does not depend on the direction of the pendulum's swing, so we may assume that the invariant plane lies in a north-south direction; then any precession would break the reflection symmetry between the northern and southern hemispheres, so the pendulum must not precess at all. (Alternatively, the Coriolis force at the equator always points vertically, and cannot torque the pendulum's invariant plane.) Now if C is made up of geodesic segments, the precession will all come from the angles where the segments meet; the total precession is equal to the net deficit angle, which in turn equals the solid angle enclosed by C modulo 2π . Finally, we can approximate any loop by a sequence of geodesic segments, so the most general result (on or off the surface of the sphere) is that the net precession is equal to the enclosed solid angle. This result may seem rather esoteric, but its generality and geometric nature suggest its depth. In fact, the mathematics describing it is essentially identical to that describing the motion of a charged particle in the field of a magnetic monopole, as well as interesting molecular, NMR, and optical systems.

A fundamental example of holonomy, involving a non-abelian symmetry, lies at the heart of general relativity. If a reference frame is parallel-transported around a closed loop in spacetime, then it is well known that the initial and final frames will not coincide. For causality's sake, we should really compare the result of parallel transport along two timelike curves with common endpoints. The final frames will be related by a Lorentz transformation, that is to say, an element of $SO(3, 1)$. The failure of the frames to coincide is an $SO(3, 1)$ holonomy. It may be used to measure the local curvature of spacetime. Thus, if we want to know the component $R_{\beta\mu\nu}^{\alpha}$ of the Riemann tensor, we should take a loop Γ in the $\mu\nu$ -plane, enclosing an infinitesimal area $dx^{\mu}dx^{\nu}$. The resulting holonomy M_{β}^{α} (an element of the Lie algebra of $SO(3, 1)$) from parallel transport around Γ will be proportional to

the area enclosed, to lowest order, and the constant of proportionality will be just the curvature component that we want:

$$M_{\beta}^{\alpha} = R_{\beta\mu\nu}^{\alpha} \cdot dx^{\mu} dx^{\nu}.$$

One last example concerns the motion of charged particles in strong magnetic fields. As is well known, in a constant and uniform magnetic field, a charged particle will move in a circle, or more precisely (in three dimensions) on a circular helix whose axis is parallel to the magnetic field direction. The motion is called cyclotron motion and the orbits cyclotron orbits, in reference to the use of such motion to guide particles at high-energy accelerators. A fundamental topic in plasma physics, with many applications in astrophysics, is how this motion is perturbed by various other effects such as inhomogeneities or time dependence in the magnetic field, electric fields, or gravitational forces. Insofar as the magnetic field is strong and reasonably homogeneous (*i.e.*, if it does not vary significantly over the radius of a cyclotron orbit), to a first approximation the motion is still cyclotron motion about the field lines, of gyrofrequency $\Omega(x) = eB(x)/mc$. The angular position of a particle relative to its guiding center axis will to lowest order be equal to the time integral of $\Omega(x(t))$. However, if the field lines are curved on a large scale, there will be both corrections to the angular position and drift corrections to the location of the guiding center.

Recently, Littlejohn ² has introduced some fresh ideas and techniques, involving geometric phases, into this old subject. This has enabled him to derive the higher-order drifts much more easily and systematically than was previously possible. In fact, even without entering into any details one can see how geometric phases enter naturally into corrections to purely cyclotronic motion. As particles rotate about their guiding centers, they are “acquiring phase”—that is, their circular motion can be parametrized by an angle that increases with time. To a first approximation, the rate of phase accumulation is the local cyclotron frequency, uniquely determined by the local value of the magnetic field. However, if our particle returns to its starting point after following a closed field line, or after drifting through some loop, the total angle through which it has turned is not just the time integral of the local frequency, but contains an additional piece depending only on the geometry of the path through the space of possible magnetic field vectors. This additional phase, which is the leading correction to the total angle, is yet another example of holonomy in a purely classical system.

We hope that the above examples, which are concrete and readily visualized, will help to put in perspective some of the more abstract applications that follow.

[1] Keith R. Symon, *Mechanics*, 2nd edition (Reading: Addison-Wesley, 1960).

- [2] Robert G. Littlejohn, "Phase anholonomy in the classical adiabatic motion of charged particles," *Phys.Rev.* **A38** (1988) 6034; "Geometry and guiding center motion," in *Contemporary Mathematics*, edited by J.E. Marsden (Providence: American Mathematical Society, 1984), Vol. 28, p.151.

The Quantum Phase, Five Years After

M. V. Berry

H.H. Wills Physics Laboratory
Tyndall Avenue, Bristol BS8 1TL, U.K.

(Received 28 April 1988.)

ABSTRACT

Classical parallel transport of vectors is described in a manner immediately generalizable to parallel transport of quantum states in parameter space. The associated anholonomy is the geometric phase. One realization of parallel transport is by adiabatic cycling of the parameters. The phase is the flux of a 2-form. The 2-form is equivalent to the antisymmetric part of a gauge-invariant quantum geometric tensor. The symmetric part of this tensor gives a natural metric on parameter space. If the parameters are themselves regarded as dynamical variables, their adiabatic dynamics are influenced by a gauge field depending on both parts of the tensor. Corrections to the geometric phase (of higher order in an adiabatic parameter) can be obtained by successive transformations to moving frames, thereby generating a renormalization map of circuits in the space of Hamiltonians; the iterates diverge in a universal way. This quantum renormalization is illustrated by classical Newtonian and Hamiltonian renormalizations for a pendulum with changing frequency. To conclude, there are some historical remarks about geometric phases.

1. Introduction

The kind invitation to write this survey article provides two welcome opportunities. First, to present the fundamentals of the subject in a new perspective, reflecting some of the many recent developments and including some new material; and second, to make some historical remarks, drawing attention to important early works and describing the genesis of my own ideas in this field.

Two concepts are crucial to the understanding of this dusty corner of quantum theory which the brooms of our understanding are beginning to disturb. They are *anholonomy* and *adiabaticity*.

Anholonomy is a geometrical phenomenon in which nonintegrability causes some variables to fail to return to their original values when others, which drive them, are altered round a cycle. The simplest anholonomy is in the parallel transport of vectors, two examples being the change in the direction of swing of a Foucault pendulum after one rotation of the earth, and the change in the direction of linear polarization of light along a twisting ray [1][2] or coiled optical fibre [3-6] whose direction is altered in a cycle. The anholonomy to be described here is quantum-mechanical, and concerns the phase of a state which is parallel-transported round a cycle [7]. Parallel transport of a quantum state will here be introduced as a simple generalization of parallel transport of a vector.

Adiabaticity is slow change and therefore denotes phenomena at the border between dynamics and statics. Adiabatic change provides the simplest (but not the only [8]) way to make quantum parallel transport happen. The variables which are cycled are parameters in the Hamiltonian of a system. If the cycling is slow, the adiabatic theorem [9] guarantees that the system returns to its original state. But it usually acquires a nontrivial phase, a manifestation of anholonomy, and this is the phenomenon of interest here.

2. Classical Parallel Transport

It is convenient to begin by obtaining the law for the ordinary parallel transport of a vector over the surface of a sphere, expressing it in a form enabling instantaneous generalization to quantum mechanics. Let the unit vector \mathbf{e} be transported by changing the unit radius vector \mathbf{r} (Fig.1) and making two demands: that $\mathbf{e} \cdot \mathbf{r}$ must remain zero and that the orthogonal triad (frame) containing \mathbf{e} and \mathbf{r} must not twist about \mathbf{r} , *i.e.*, $\boldsymbol{\Omega} \cdot \mathbf{r} = 0$ where $\boldsymbol{\Omega}$ is the angular velocity of the triad. These conditions define parallel transport of \mathbf{e} and lead to the law

$$\dot{\mathbf{e}} = \boldsymbol{\Omega} \wedge \mathbf{e} \quad \text{where} \quad \boldsymbol{\Omega} = \mathbf{r} \wedge \dot{\mathbf{r}} \quad (1)$$

This law is nonintegrable; when \mathbf{r} returns to its original direction after a circuit C on the sphere, \mathbf{e} does not return (in spite of never having been

twisted) but has turned through an angle $\alpha(C)$ which is the anholonomy now to be determined. Define $\mathbf{e}' \equiv \mathbf{r} \wedge \mathbf{e}$ (so that \mathbf{r} , \mathbf{e} , \mathbf{e}' form an orthogonal triad) and the complex unit vector

$$\boldsymbol{\psi} \equiv (\mathbf{e} + i\mathbf{e}')/\sqrt{2} \quad (2)$$

in the plane perpendicular to \mathbf{r} . In terms of $\boldsymbol{\psi}$, the parallel transport law (1) (which holds for \mathbf{e}' as well as \mathbf{e}) takes the simple form

$$\text{Im } \boldsymbol{\psi}^* \dot{\boldsymbol{\psi}} = 0 \quad \text{i.e.,} \quad \text{Im } \boldsymbol{\psi}^* d\boldsymbol{\psi} = 0 \quad (3)$$

where $d\boldsymbol{\psi}$ is the change in $\boldsymbol{\psi}$ resulting from a change $d\mathbf{r}$.

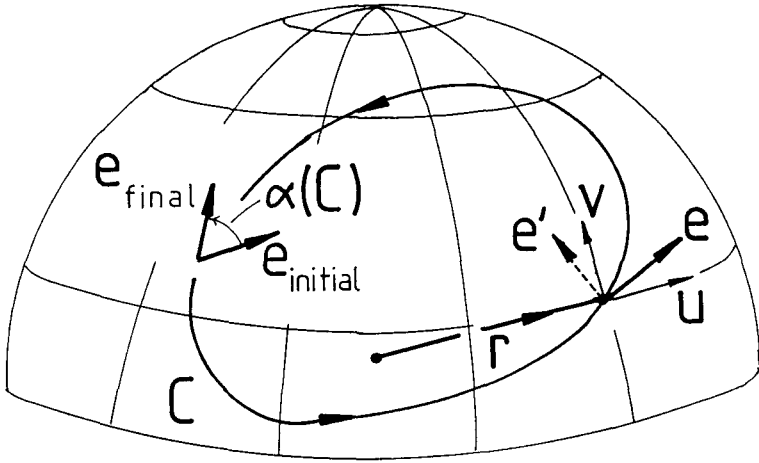


Figure 1. Rotation by $\alpha(C)$ after parallel transport of vector \mathbf{e} round circuit C on a sphere.

To find $\alpha(C)$ we chart the passage of \mathbf{e} and \mathbf{e}' relative to a local basis of unit vectors $\mathbf{u}(\mathbf{r}), \mathbf{v}(\mathbf{r})$ (Fig.1) defined at each point on the sphere. For example, we could choose \mathbf{u} and \mathbf{v} to lie along the parallel of latitude θ and meridian of longitude ϕ at $\mathbf{r} = (\sin\theta \cos\phi, \sin\theta \sin\phi, \cos\theta)$, i.e.,

$$\mathbf{u} = (-\sin\phi, \cos\phi, 0), \quad \mathbf{v} = (-\cos\theta \cos\phi, -\cos\theta \sin\phi, \sin\theta). \quad (4)$$

Specifying a local basis is equivalent to specifying the complex unit vector

$$\mathbf{n}(\mathbf{r}) \equiv (\mathbf{u}(\mathbf{r}) + i\mathbf{v}(\mathbf{r}))/\sqrt{2} \quad (5)$$

If the angle between the transported \mathbf{e} and the local \mathbf{u} is $\alpha(t)$, (2) and (5) give

$$\boldsymbol{\psi} = \mathbf{n} \exp(-i\alpha) \quad (6)$$

whence (3) gives the anholonomy as

$$\begin{aligned} \alpha(C) &= \oint d\alpha = \text{Im} \oint \mathbf{n}^* \wedge \cdot d\mathbf{n} \\ &= \text{Im} \iint_{\partial S=C} d\mathbf{n}^* \cdot d\mathbf{n} \end{aligned} \quad (7)$$

where in the last equality Stokes' theorem has been used and the integral is over the area on the sphere bounded by C . It is an important result that the integrand in (7) is independent of the choice of local basis \mathbf{u}, \mathbf{v} : a change in this choice can be represented by a rotation $\mu(\mathbf{r})$ which induces the gauge transformation

$$\mathbf{n}(\mathbf{r}) \rightarrow \mathbf{n}'(\mathbf{r}) \exp\{i\mu(\mathbf{r})\} \quad (8)$$

under which $d\mathbf{n}^* \wedge \cdot d\mathbf{n}$ is invariant.

In terms of arbitrary parameters X_1, X_2 specifying \mathbf{r} (i.e., position on the sphere), Eq. (7) can be written explicitly as

$$\alpha(C) = \text{Im} \iint_{\partial S=C} dX_1 dX_2 (\partial_1 \mathbf{n}^* \cdot \partial_2 \mathbf{n} - \partial_2 \mathbf{n}^* \cdot \partial_1 \mathbf{n}) \quad (9)$$

where ∂_j denotes $\partial/\partial X_j$. The choice $X_1 = \theta, X_2 = \phi$, together with (4), yields the integrand $d\theta d\phi \sin\theta$, which is simply the area element on the sphere, leading to the old result that the anholonomy $\alpha(C)$ is the *solid angle* subtended by C at the centre of the sphere.

3. Quantum Parallel Transport

To make the generalization to quantum mechanics, we replace the complex unit vector $\boldsymbol{\psi}$ by a normalized quantum state $|\psi\rangle$, i.e., a unit vector in a Hilbert space, and position $\mathbf{r} = (X_1, X_2)$ on the sphere by position $X = (X_1, X_2, \dots)$ in a space of parameters governing the physical system represented by $|\psi\rangle$. At each X , $|\psi\rangle$ is defined up to a phase (just as \mathbf{e} was defined up to a rotation at each \mathbf{r}). Then a natural transport law [10] governing the phase of $|\psi\rangle$ as X varies is provided by reinterpreting (3) as the connection

$$\text{Im} \langle \psi | d\psi \rangle = 0. \quad (10)$$

Like (3), this law is nonintegrable: when X is taken round a circuit C , $|\psi\rangle$ returns with a changed phase. This change is the *quantum geometric phase* $\gamma(C)$; thus

$$\langle \psi_{\text{initial}} | \psi_{\text{final}} \rangle = \exp\{i\gamma(C)\}. \quad (11)$$

To find γ we again introduce a local basis by choosing at each X a definite (and so of course single-valued) state $|n(X)\rangle$, relative to which $|\psi\rangle$ is defined by

$$|\psi\rangle = |n(X)\rangle \exp(i\gamma) \quad (12)$$

Then (10) gives

$$\begin{aligned} \gamma(C) &= \oint d\gamma = -\text{Im} \oint \langle n|dn\rangle \\ &= -\text{Im} \iint_{\partial S=C} \langle dn|\wedge|dn\rangle \equiv -\iint_{\partial S=C} V(X). \end{aligned} \quad (13)$$

The integrand $V = \text{Im} \langle dn|\wedge|dn\rangle$ is the *phase 2-form*, whose flux through C gives the geometric phase. V is invariant under the gauge transformation

$$|n(X)\rangle \rightarrow |n'(X)\rangle \equiv |n(X)\rangle \exp\{i\mu(X)\} \quad (14)$$

For this mathematics to represent physics, it must be possible to implement the connection (10) by the Schrodinger equation

$$i\hbar|\dot{\Psi}\rangle = \hat{H}|\Psi\rangle \quad (15)$$

governing the evolution of any state $|\Psi\rangle$. A simple way [7] is to incorporate the parameters X into the Hamiltonian and change them slowly. Then the adiabatic theorem guarantees that in the absence of degeneracies (a restriction that can be removed [46]) $|\Psi\rangle$ will cling to one of the eigenstates of $\hat{H}(X(t))$, defined by

$$\hat{H}(X)|\psi\rangle = E_n(X)|\psi\rangle \quad (16)$$

The adiabatic ansatz

$$|\Psi\rangle \approx |\psi\rangle \exp\left\{-\frac{i}{\hbar} \int_0^t dt' E_n(X(t'))\right\} \quad (17)$$

then gives the connection (10) immediately upon projecting (15) onto $|\psi\rangle$. The state $|n(X)\rangle$ in the 2-form (13) is any solution of (16) with a definite phase at each X .

Because of (17), the total phase change of $|\Psi\rangle$ includes a dynamical part as well as the $\gamma(C)$ being studied here. Thus

$$\langle \Psi_{\text{final}}|\Psi_{\text{initial}}\rangle = \exp\{i(\gamma_d + \gamma C)\} \quad (18)$$

where, for a circuit that takes a time T ,

$$\gamma_d = -\frac{1}{\hbar} \int_0^T dt E_n(X(t)) \quad (19)$$

One might say that γ_d and $\gamma(C)$ give the system's best answers to two questions about its adiabatic circuit. For γ_d the question is: how long did your journey take? For $\gamma(C)$ it is: where did you go?

Aharonov and Anandan [8] give a different interpretation of parallel transport. They regard the parameters X as labelling the state, rather than \hat{H} , so that X_1, X_2, \dots are coordinates in the *projective Hilbert space* that includes all quantum states, but where states differing only in phase (or normalization) are represented by the same point. Then a state $|\Psi\rangle$ evolving under Eq. (15) (not necessarily adiabatically) so as to return in T to the same X acquires a phase (18), with geometric part (13) (where the phase of $|n(X)\rangle$ is an arbitrary function of X) and dynamical part given by

$$\gamma_d = -\frac{1}{\hbar} \int_0^T dt \langle \Psi | \hat{H} | \Psi \rangle \quad (20)$$

instead of Eq. (19). The relation between the two approaches is that in the adiabatic case X parameterizes that part of the projective Hilbert space corresponding to the n th eigenstate of the chosen family of Hamiltonians $\hat{H}(X)$.

Several experiments have measured the geometric phase for particles, with spin 1/2 (neutrons [11]), spin 1 (photons [3]) and spin 3/2 (chlorine nuclei [12]). These depend on the result [7] that when \hat{H} is a rotationally symmetric function of the spin, *i.e.*,

$$\hat{H} = F(\boldsymbol{\sigma} \cdot \mathbf{X}) \quad (21)$$

where $\mathbf{X} = (X_1, X_2, X_3)$ and $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \sigma_3)$ is the vector spin operator, the geometric phase for the state with spin component n along \mathbf{X} is

$$\gamma_n(C) = -n \Omega(C) \quad (22)$$

where $\Omega(C)$ is the solid angle subtended by C at $\mathbf{X} = 0$.

These experiments all employ a superposition of eigenstates, rather than a single one, so that

$$\begin{aligned} |\Psi_{\text{initial}}\rangle &= \sum_n a_n |n\rangle \\ |\Psi_{\text{final}}\rangle &= \sum_n a_n |n\rangle \exp\{i(\gamma_{dn} + \gamma_n(C))\} \end{aligned} \quad (23)$$

At the end, that is after X has been cycled, an observable \hat{A} , which does not commute with the final \hat{H} , is measured (for example with a polarizer). Thus

$$\begin{aligned} \langle \hat{A} \rangle &= \sum_n |a_n|^2 \langle n | \hat{A} | n \rangle + 2 \operatorname{Re} \sum_{m \neq n} a_n^* a_m \langle m | \hat{A} | n \rangle \\ &\quad \times \cos \left\{ [\gamma_{dn} + \gamma_n(C)] - [\gamma_{dm} + \gamma_m(C)] \right\}. \end{aligned} \quad (24)$$

The oscillatory terms reveal $\gamma_n(C)$. This scheme has proved more convenient than the earlier suggestion [7] of splitting an ensemble of systems (*e.g.*, a beam of particles) into two subensembles, one being driven by an \hat{H} which is cycled and the other by an \hat{H} which is not, and then recombining the subensembles to detect $\gamma(C)$ by interference. (That is, instead of using one state and two Hamiltonians it is preferable to use two states — at least — and one Hamiltonian.)

Hannay [13] found an analogue of the geometric phase for *classical* systems. This was based on the simple observation that a quantum system in an eigenstate is an oscillator (because of the time factor $\exp(-iE_n t/\hbar)$), so that classical oscillators should exhibit similar anholonomy when parameters that govern them are cycled. The phase is now an angle, which may be an angle in space, like that of a wheel, or — more commonly — an abstract angle variable in phase space as with a harmonic oscillator. If the classical system is multiply periodic (integrable) for all X , with N freedoms (that is, coordinates $\mathbf{q} = (q_1 \cdots q_N)$ and momenta $\mathbf{p} = (p_1 \cdots p_N)$ and Hamiltonian $H(\mathbf{q}, \mathbf{p}; X)$, its orbit for fixed X winds round an N -torus [14] in phase space, with N angle variables $\boldsymbol{\theta} = (\theta_1 \cdots \theta_N)$ increasing uniformly. Conjugate to $\boldsymbol{\theta}$ are N adiabatically conserved actions $\mathbf{I} = (I_1 \cdots I_N)$ which label the torus. After a slow cycle of X the angles have acquired shifts which contain a geometric as well as a dynamical part. For a spinning particle [15-17] this classical anholonomy is the angle shift given by ordinary parallel transport of a vector.

Underlying Hannay's angles is a *classical 2-form*. This is the classical limit of the phase 2-form in Eq. (13), and semiclassical asymptotics [18] provides the expression

$$V(X) \xrightarrow{\hbar \rightarrow 0} - \langle d\mathbf{p} \wedge \cdot d\mathbf{q} \rangle / \hbar \quad (25)$$

whose symbols should be interpreted as follows. The wedge product \wedge links the d 's in parameter space. The scalar product \cdot links \mathbf{p} and \mathbf{q} . $\langle \rangle$ denotes an average over the angles on the torus labelled \mathbf{I} which at X corresponds [19] to the quantum state $|n\rangle$, *i.e.*, $\langle \rangle = \int_0^{2\pi} d\theta_1 \cdots d\theta_N / (2\pi)^N$. $d\mathbf{q}$ is the coordinate displacement linking corresponding points (labelled by the same $\boldsymbol{\theta}$) on the tori \mathbf{I} at X and $X + dX$, and similarly for $d\mathbf{p}$.

It is amusing to note that if the $2N$ variables \mathbf{q} and \mathbf{p} are replaced by the N complex variables

$$\mathbf{n} = (n_1 \cdots n_N) \equiv (\mathbf{q} + i\mathbf{p}) / \sqrt{2\hbar} \quad (26)$$

then (25) takes the form

$$V(X) \xrightarrow{\hbar \rightarrow 0} \text{Im} \langle d\mathbf{n}^* \wedge \cdot d\mathbf{n} \rangle \quad (27)$$

which bears a close formal resemblance both to the quantum expression (13) and the geometrical formula (7).

If the classical motion is not multiply periodic, that is if it is wholly or partly chaotic, the question of the classical limit of V is more delicate. It is tempting to claim that the limit is (25) for nonintegrable as well as integrable motion, but it is difficult to interpret the average $\langle \rangle$ and the displacements $d\mathbf{q}$ and $d\mathbf{p}$. In one of several interpretations, obtained by a semiclassical argument (not yet published) in collaboration with M. Wilkinson, $\langle \rangle$ denotes a time average over all points on an infinite orbit, and $d\mathbf{q}$ and $d\mathbf{p}$ link simultaneous points on the orbits for X and $X + dX$. For nonintegrable systems, however, it is not easy to express this result by replacing $\langle \rangle$ by a phase-space integral over the manifold explored by the orbit, because it is not clear what are then the ‘corresponding points’, linked by $d\mathbf{q}$ and $d\mathbf{p}$, on the manifolds for X and $X + dX$ (for an ergodic system these are the two constant-energy surfaces).

4. The Quantum Geometric Tensor

The central mathematical object underlying the quantum phase is the 2-form $V = \text{Im} \langle dn \wedge dn \rangle$. This is equivalent to an antisymmetric second-rank tensor field $V_{ij}(X)$ on the parameter space (or projective Hilbert space) with a quantum state $|n(X)\rangle$ defined at each point, namely

$$V_{ij}(X) = \text{Im}\{\langle \partial_i n | \partial_j n \rangle - \langle \partial_j n | \partial_i n \rangle\} \quad (28)$$

This tensor is invariant under the gauge transformation (14), but it is not the only such invariant tensor. More general is the *quantum geometric tensor*

$$T_{ij}(X) \equiv \langle \partial_i n | (1 - |n\rangle\langle n|) | \partial_j n \rangle \quad (29)$$

which is Hermitian, *i.e.*, $T_{ij} = T_{ji}^*$. The projector $|n\rangle\langle n|$ is essential to the gauge invariance. The imaginary part of T_{ij} is simply $V_{ij}/2$, so we can write

$$T_{ij} = g_{ij} + iV_{ij}/2 \quad (30)$$

where g_{ij} is the real symmetric tensor field $\text{Re} T_{ij}$.

We know the quantum meaning of V_{ij} : its flux gives the phase $\gamma(C)$. Therefore, it is natural to ask whether g_{ij} has significance. The answer is that g_{ij} provides a natural means of measuring distances along paths in parameter space; it is the *quantum metric tensor*. To understand why, observe that a natural measure of the squared distance between two nearby quantum states is the deviation from unity of their scalar product. If the states are $|1\rangle$ and $|2\rangle$ this gives, for the distance between the corresponding points X_1 and X_2 in parameter space,

$$\Delta s_{12}^2 = 1 - |\langle 1|2\rangle|^2 \quad (31)$$

Taking the limit $1 \rightarrow 2$, and using the fact that all states are normalized, we obtain (using the summation convention for repeated indices i and j)

$$\begin{aligned} ds^2 &= \langle dn | (1 - |n\rangle\langle n|) |dn\rangle = \langle \partial_i n | (1 - |n\rangle\langle n|) | \partial_j n \rangle dX_i dX_j \\ &= T_{ij} dX_i dX_j = g_{ij} dX_i dX_j \end{aligned} \quad (32)$$

as claimed. The quantum tensor was introduced in an interesting paper by Provost and Vaille [50].

From its structure, g_{ij} can never give a negative ds^2 : in fact it is a positive semidefinite metric. Along a finite path (not necessarily closed) between $|1\rangle$ and $|2\rangle$, the quantum distance is

$$s_{12}(C) = \int_1^2 (g_{ij} dX_i dX_j)^{1/2}. \quad (33)$$

Page [33] and Bouchiat and Gibbons [41] give explicit forms for some metrics on the full Hilbert and projective Hilbert spaces.

The simplest example is a 2-state system, for which \hat{H} has the form (21), with $\hat{\sigma}$ the 3 Pauli matrices. If we take X as a unit vector, specified by parameters θ, ϕ (polar angles), the eigenstates are

$$|+\rangle = \begin{pmatrix} \cos(\theta/2) e^{i\phi/2} \\ \sin(\theta/2) e^{-i\phi/2} \end{pmatrix}, \quad |-\rangle = \begin{pmatrix} \sin(\theta/2) e^{i\phi/2} \\ -\cos(\theta/2) e^{-i\phi/2} \end{pmatrix} \quad (34)$$

For both of these, (32) gives $ds^2 = d\theta^2 + \sin^2\theta d\phi^2$, and this is the natural metric on the sphere of parameters (which in this case is also the projective Hilbert space).

Some interesting questions are suggested by this identification of g_{ij} as a metric on parameter space:

- (i) Do the geodesics, and in particular the shortest paths, connecting non-neighbouring states $|1\rangle$ and $|2\rangle$ have physical significance? One possibility, suggested by the work of Pancharatnam [20][21], is that the geodesics are the special paths along which the state preserves its phase in the sense that $\langle 1|2\rangle$ is real. This is true for the 2-state system just discussed, but seems to fail otherwise (probably for reasons of codimension). It is worth remarking that as $2 \rightarrow 1$ the overlap $\langle 1|2\rangle$ is real to second as well as first order in dX , for any path whatever.
- (ii) Can the geodesics be chaotic? This would require parameters X and states $|n(X)\rangle$ for which the Riemann curvature defined in terms of g_{ij} is negative (at least in some places) and the space is compact.
- (iii) Do *families* of geodesics (for example those issuing in different directions from the same point) exhibit the generic caustic singularities classified by catastrophe theory [22][23]? Do any such caustics have physical meaning? In 2-state systems the geodesics from X focus nongenerically at the

antipodal point on the sphere, where the state is orthogonal to $|n(X)\rangle$, but again this appears to be a special situation.

- (iv) Is there any meaning or interest in *quantizing* the geodesic motion in parameter space, for example by taking as Hamiltonian the Laplace-Beltrami operator $g^{-1/2}\partial_i g^{-1/2}g_{ij}\partial_j$ (where $g \equiv \det g_{ij}$)? Such quantizations are different from that described in the next section.

5. Dynamics of the Parameters

Until now we have regarded X as classical parameters which can be altered arbitrarily and which are unaffected by the quantum system they drive. But no physical action is unilateral and in reality X are themselves dynamical variables of a ‘heavy’ system coupled to the ‘light’ system (what we have so far called ‘the’ system) and therefore subject to reaction from it. Indeed the earliest application of the adiabatic theorem was the Born-Oppenheimer theory of molecules, in which X are coordinates describing the positions of the (heavy) nuclei and the light system is the electrons. Recently it has been pointed out [24–27] that in lowest order the reaction of the light system on the heavy dynamics is through a gauge field consisting of a vector potential whose curl is the phase 2-form V , and a scalar potential. Here I will show that what the gauge field really depends on is the quantum geometric tensor T_{ij} of section 3.

Let the heavy momenta, conjugate to X_i , be P_i . Then a fairly general nonrelativistic quantum Hamiltonian for the coupled system is

$$\hat{H}_{\text{tot}} = \frac{1}{2} \sum_{ij} Q_{ij} \hat{P}_i \hat{P}_j + H(\hat{\xi}; \hat{X}), \quad (35)$$

in which Q_{ij} is an inverse mass tensor, $\hat{\xi}$ are the dynamical variables of the light system (coordinates, momenta, spins, . . .) and H our previous Hamiltonian in which the X were regarded as parameters and which has eigenstates $|n(X)\rangle$ and energies $E_n(X)$. In the position representation for the heavy system, that is $\hat{P}_i = -i\hbar\partial_i$, the adiabatic ansatz is to write the full quantum state in the separated form

$$\langle X|\Psi\rangle \approx \Psi_{\text{heavy}}(X)|n(X)\rangle \quad (36)$$

and to consider the effective Hamiltonian governing Ψ_{heavy} to be

$$\hat{H}_{\text{eff}} = \langle n(X)|\hat{H}_{\text{tot}}|n(X)\rangle. \quad (37)$$

In \hat{H}_{eff} the reaction of the light on the heavy system comes from the action of the gradient operators \hat{P}_i on the X -dependence of $|n\rangle$. A straightforward calculation gives

$$\hat{H}_{\text{eff}} = \frac{1}{2} \sum_{ij} Q_{ij} \left\{ \hat{P}_i - A_i(\hat{X}) \right\} \left\{ \hat{P}_j - A_j(\hat{X}) \right\} + \Phi(\hat{X}) + E_n(\hat{X}) \quad (38)$$

where

$$A_i(X) = i\hbar \langle n | \partial_i n \rangle \quad (39)$$

and

$$\Phi(X) = \frac{\hbar^2}{2} \sum_{ij} Q_{ij} g_{ij}(X) \quad (40)$$

Here the emphasis is on the gauge potentials Φ and A_i — the scalar $E_n(X)$ is the ‘potential surface’ studied in conventional Born-Oppenheimer theory. Although (38) is a quantum Hamiltonian it can be used in suitable circumstances to calculate the *classical* motion of the heavy system, which will be affected by the fields A_i and Φ .

The physical effects of the vector potential A_i depend only on the ‘magnetic’ field

$$F_{ij} = \partial_i A_j - \partial_j A_i = -\hbar V_{ij} \quad (41)$$

(including its singularities and values in inaccessible regions — I am not denying the Aharonov-Bohm effect for heavy systems!). Thus the ‘magnetic’ field seen by the heavy system is the antisymmetric part of the quantum geometric tensor. The symmetric part of T_{ij} determines the ‘electric’ potential via Eq. (40). For an isotropic mass tensor, *i.e.*, $Q_{ij} = \delta_{ij}/M$, Φ depends on $Tr g_{ij}$. It is a curious asymmetry that the ‘electric’ field depends on the *gradients* of g_{ij} whereas the ‘magnetic’ field depends on V_{ij} itself.

The singularities of the gauge field are the *degeneracies* X^* of the spectrum, where $E_n(X^*) = E_{n\pm 1}(X^*)$. It is already known [7] that the ‘magnetic’ field V_{ij} (2-form) has monopole singularities. From the definition (29) of T_{ij} it is clear that g_{ij} has similar singularities, so that the ‘electric’ field near X^* is an *inverse-cube* force.

The situation near a degeneracy can be described by a special case of a simple model, which is of independent interest (and which has been studied from a different viewpoint by Anandan and Aharonov [28]), where the spin s of one (light) particle is coupled to the spatial coordinates of a second otherwise free (heavy) particle. Thus

$$\hat{H}_{\text{tot}} = \frac{1}{2M} \hat{P}^2 + F(\hat{\mathbf{X}} \cdot \hat{\boldsymbol{\sigma}}) \quad (42)$$

Near a degeneracy the appropriate model is a 2-state light system, so that we should take $s = \frac{1}{2}$, with linear coupling $F \propto \mathbf{X} \cdot \boldsymbol{\sigma}$.

The eigenvalues of $\mathbf{X} \cdot \hat{\boldsymbol{\sigma}}$ are nX , where $X \equiv |\mathbf{X}|$ and $-s \leq n \leq s$. The quantum tensor for the state $|n\rangle$ can be shown to be

$$T_{ij}^n(X) = \frac{1}{2X^2} \left\{ \left(s(s+1) - n^2 \right) \left(\mathbf{e}_i \cdot \mathbf{e}_j - (\mathbf{e}_i \cdot \mathbf{x})(\mathbf{e}_j \cdot \mathbf{x}) \right) \mp in(\mathbf{e}_i \wedge \mathbf{e}_j) \cdot \mathbf{x} \right\} \quad (43)$$

where $\mathbf{x} = \mathbf{X}/|\mathbf{X}|$ and \mathbf{e}_i is the unit vector along the i direction. The metric tensor g_{ij} has a zero eigenvalue, corresponding to radial parameter

displacements, which simply scale H leaving the states $|n\rangle$ unaffected: radial motions cover zero distance.

From Eqs. (38)–(40), the *classical* Newtonian equation for the heavy particle involves the Lorentz force from the magnetic monopole and the ‘electric’ force

$$-\nabla_{\mathbf{X}}\Phi(\mathbf{X}) = -\frac{\hbar^2}{2M}\nabla_{\mathbf{X}}\text{Tr}g_{ij} = \frac{\hbar^2(s(s+1) - n^2)}{MX^3}\mathbf{x} \quad (44)$$

This is of centrifugal type, and repels the parameters from a degeneracy (becoming significant at a distance of order $M^{-1/3}$), thereby tending to preserve the validity of the adiabatic approximation. We obtain, when the light particle is in the n th spin state,

$$M\ddot{\mathbf{X}} = \frac{S_z}{2X^3}\dot{\mathbf{X}}\wedge\mathbf{X} + \frac{(S^2 - S_z^2)}{MX^4}\mathbf{X} - \frac{nF'(nX)}{X}\mathbf{X} \quad (45)$$

where $S_z \equiv n\hbar$ and $S^2 \equiv \hbar^2s(s+1)$. This describes integrable motion, with conserved energy and modified angular momentum $M\mathbf{X}\wedge\dot{\mathbf{X}} - S_z\mathbf{X}/X$.

6. Adiabatic Renormalization

Now we return to the adiabatic scenario of section 3 and realize that γ_d and $\gamma(C)$ in Eq. (18) are but the first two terms in an infinite series involving powers of an adiabatic slowness parameter ϵ , influencing the dynamics through \hat{H} whose time-dependence enters in the combination ϵt . The dominant term is γ_d (Eq. 19) and is of order ϵ^{-1} . The next term is $\gamma(C)$, whose unique feature — and the reason for its being called geometric — is that it is independent of ϵ , and so depends only on the sequence of Hamiltonians along the circuit and not on its time history.

This uniqueness is not threatened by the observation that transformation to a moving frame (a common practice in problems involving spin [11]) can make $\gamma(C)$ appear ‘dynamical’ by making it emerge from a correction to the energy rather than as anholonomy: the geometric structure of $\gamma(C)$ is independent of how it is derived.

Transformations to moving frames have however another interest, in that they form the basis of a renormalization (iteration) technique for generating higher-order corrections to the phase. Details of the technique have been published elsewhere [29]; here I will outline the central idea, and give an example.

Let the Hamiltonian $\hat{H}_0(t)$ generating the quantum motion be cyclic, in the sense that $\hat{H}_0(+\infty) = \hat{H}_0(-\infty)$, and let it have instantaneous eigenstates $|n_0(t)\rangle$ and energies $E_0(n, t)$. The evolving state $|\Psi_0(t)\rangle$ is determined by

$$i|\dot{\Psi}_0(t)\rangle = \hat{H}_0(t)|\Psi_0(t)\rangle \quad (46)$$

with the initial condition

$$|\Psi_0(-\infty)\rangle = |n_0(-\infty)\rangle \equiv |N\rangle \quad (47)$$

After the cycle, *i.e.*, at $t = +\infty$, $|\Psi_0\rangle$ will have returned only approximately to $|N\rangle$, so a phase can be defined precisely by

$$\gamma \equiv \text{Im} \log \langle N | \Psi_0(+\infty) \rangle - \gamma_d \quad (48)$$

The geometric phase $\gamma(C)$ (Eq. 13) is $\lim_{\epsilon \rightarrow 0} \gamma$. The aim is to obtain increasingly accurate approximations to $\gamma - \gamma(C)$. It is worth emphasizing that the non-aim is the determination of the nonadiabatic transition probability $1 - |\langle N | \Psi_0(+\infty) \rangle|^2$, because this is the usual objective of adiabatic theory, and that the non-method is perturbation theory, because this is the usual technique [30][49].

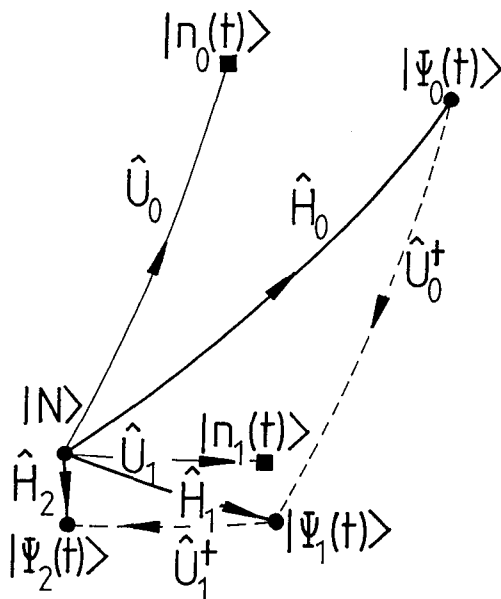


Figure 2. Renormalization in Hilbert space.

To explain the method used instead, we refer to Fig. 2. When ϵ is small we expect $|\Psi_0(t)\rangle$ to be close to $|n_0(t)\rangle$. This suggests that defining a unitary transformation $\hat{U}_0(t)$ by

$$|n_0(t)\rangle = \hat{U}_0(t)|N\rangle \quad (49)$$

will be useful. The inverse operator U_0^\dagger sends $|n_0(t)\rangle$ back to $|N\rangle$, that is, it freezes the moving eigenstate. Therefore \hat{U}_0^\dagger should almost freeze the evolving state $|\Psi_0(t)\rangle$, and so we define

$$|\Psi_1(t)\rangle \equiv \hat{U}_0^\dagger |\Psi_0(t)\rangle. \quad (50)$$

We are attempting to follow $|\Psi_0(t)\rangle$ by transforming to a moving frame. The Hamiltonian governing $|\Psi_1\rangle$ is

$$\hat{H}_1 = \hat{U}_0^\dagger \hat{H}_0 \hat{U}_0 - i \hat{U}_0^\dagger \dot{\hat{U}}_0, \quad (51)$$

in which the second term is the quantum analogue of the inertial forces generated classically by transforming to a moving frame.

Now that the original problem has been reduced to one of the same form but involving $|\Psi_1\rangle$ and \hat{H}_1 instead of $|\Psi_0\rangle$ and \hat{H}_0 , it is natural to iterate the process by defining $|\Psi_2\rangle \equiv \hat{U}_1^\dagger |\Psi_1\rangle$, where \hat{U}_1^\dagger freezes the eigenstates $|n_1\rangle$ of \hat{H}_1 . This defines a *renormalization map* $\hat{H}_k \rightarrow \hat{H}_{k+1}$ in Hamiltonian space. The form of the map is simple when written in a basis of initial states (which are unaffected by renormalization) and with the phases of the eigenstates chosen so that they are parallel-transported, *i.e.*, $\langle n_k | \dot{n}_k \rangle = 0$:

$$\langle M | \hat{H}_{k+1} | N \rangle = E_k(n, t) \delta_{MN} - i \frac{\langle m_k(t) | \dot{\hat{H}}_k(t) | n_k(t) \rangle}{E_k(m, t) - E_k(n, t)} (1 - \delta_{MN}) \quad (52)$$

The k th approximant $\gamma^{(k)}$ to the phase is obtained by neglecting the off-diagonal terms in \hat{H}_{k+1} . $\gamma^{(k)}$ is the sum of the phase anholonomies of the Hamiltonians $\hat{H}_0 \dots \hat{H}_k$ (arising from the continuation of $|n_k(t)\rangle$ from $t = -\infty$ to $t = +\infty$ and reflected as phase factors $\langle N | U_k(+\infty) | N \rangle$), together with an additional term involving E_k [29]. (A contrary choice of phases, *i.e.*, $|n_k(+\infty)\rangle = |n_k(-\infty)\rangle$, gives $\langle N | U_k(+\infty) | N \rangle = 1$, but now the diagonal terms in Eq. (52) contain extra terms $-i \langle n_k | \dot{n}_k \rangle$ and all corrections — including $\gamma^{(0)} = \gamma(C)$ as mentioned previously — appear dynamical.)

Each renormalization produces a new Hamiltonian which over $-\infty < t < +\infty$ traverses a loop in Hamiltonian space. If the renormalizations converged, successive loops would get smaller (by a factor ϵ each time). But this does not, and indeed cannot, happen. If it did, $\langle \Psi(-\infty) | \Psi(+\infty) \rangle$ would have modulus unity, contradicting the existence of transitions to other states. The accumulation of inertial forces in successive renormalizations defeats our attempts to follow the motion, which slips out of control, causing the scheme to diverge.

Nevertheless, the corrections generated by renormalization do get smaller at first, and enable γ to be determined with an error of order $\exp(-1/\epsilon)$, which occurs after $k \sim 1/\epsilon$ renormalizations. A detailed exploration [29] of 2-state systems (the simplest nontrivial case, for which the geometry of the loop map can be made explicit) reveals that the Hamiltonian loops (which lie on a 2-sphere) get smaller and then larger in a universal way (that is, almost always independent of the form of the initial loop).

This procedure is typical of asymptotic procedures and occurs also in the more usual adiabatic perturbation theory. It prompts interesting questions. What is the dynamical significance of the moving frame that produces the best approximant to γ , generated by $\hat{U}_{k \sim 1/\epsilon} \hat{U}_{k-1} \dots \hat{U}_0$? Can the exponential

residue $\gamma - \gamma^{(k)}$ be more closely approximated by generalizing the Borel (or some other) resummation method [47]?

It is instructive to illustrate adiabatic renormalization with the *classical* problem which gave birth to the entire subject, namely the Ehrenfest-Einstein pendulum [31] whose frequency is slowly changed. Newton's equation is

$$\partial_t^2 x(t) + \omega^2(t) x(t) = 0 \quad (53)$$

in which the frequency $\omega(t)$ is a smooth nonzero function with $\omega(+\infty) = \omega(-\infty) \equiv \omega_\infty$. The same equation describes the (time-independent) quantum mechanics of a beam of particles with energy E encountering a potential well or hill $V(x)$ such that $E > V(x)$ for all x .

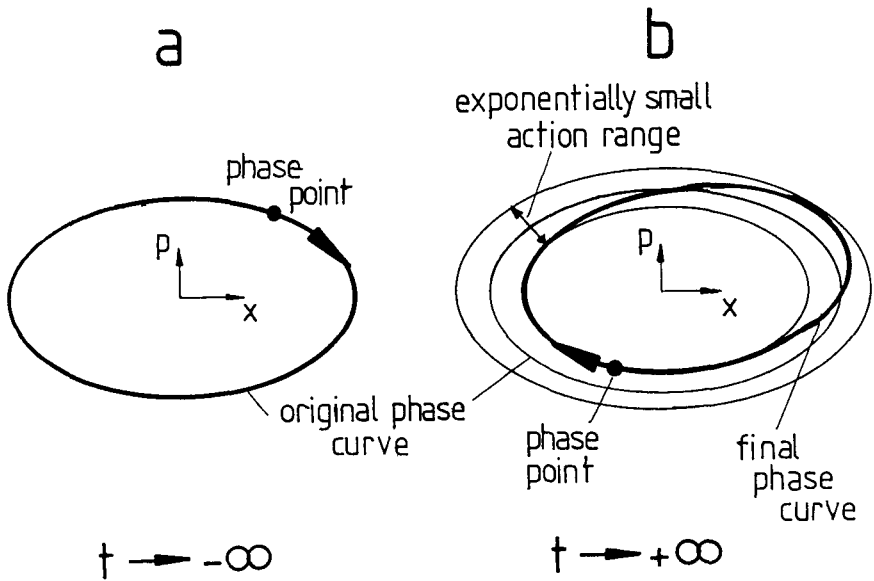


Figure 3. (a) Initial and (b) final phase portraits for slowly-altered pendulum.

Consider motion in the phase plane with variables x and $p = \dot{x}$. Initially, *i.e.*, as $t \rightarrow -\infty$, each phase point moves round an ellipse with frequency ω_∞ (Fig. 3a). The subsequent motion lies on a curve that at each instant approximates one of the elliptical contours of the Hamiltonian

$$H(x, p, t) = \frac{1}{2} [p^2 + \omega^2(t)x^2] \quad (54)$$

at that time. These subsequent ellipses have approximately the same area as the original one, because the adiabatically-conserved action is $area/2\pi$.

As $t \rightarrow +\infty$, then, the phase point is close to its original ellipse and we can ask: where is it on the ellipse, *i.e.*, what is its phase?

This would be a question about Hannay's angle were it not for the fact that this is a classical problem without anholonomy, so that the angle we seek consists entirely of nonadiabatic corrections. By identifying the solution of (53) with quantum transmitted and reflected waves, it can be shown that the oscillation which begins as

$$x + ip/\omega_\infty = A \exp \left\{ -i \left(\int_0^t dt' \omega(t') + \sigma \right) \right\} \quad (t \rightarrow -\infty) \quad (55)$$

ends as

$$x + ip/\omega_\infty = A \exp \left\{ -i \left(\int_0^t dt' \omega(t') + \sigma \right) \right\} [T^{-1} + RT^{-1} \exp(2i\sigma)] \quad (t \rightarrow +\infty) \quad (56)$$

where A is a real constant and R and T are the complex quantal reflection and transmission coefficients.

Therefore the phase shift depends on the initial phase σ , but this dependence is slight because R is exponentially small in the slowness parameter ϵ (if ω depends on ϵt). In any case, we can define a phase by averaging over σ , with the exact result

$$\gamma \equiv -\frac{1}{2\pi} \int_{-\pi}^{\pi} d\sigma \lim_{t \rightarrow \infty} \left[\text{Im} \log (x + ip/\omega_\infty) + \int_0^t dt' \omega(t') \right] = \text{Im} \log T. \quad (57)$$

Thus 'Hannay's angle' is here the phase of the transmission coefficient. The final action I also depends on σ , but the range is $(I_{\max} - I_{\min})/I_{\text{initial}} = 4|R|/|T|^2$ which again is of order $\exp(-1/\epsilon)$; the whole initial ellipse of phase points evolves ultimately into one exponentially close to it and deforming periodically with frequency ω_∞ (Fig. 3b). Newtonian renormalization of (53) is based on the transformation

$$x(t) \equiv \frac{x_1(t_1)}{\omega^{1/2}(t)}; \quad t_1 \equiv \int_0^t dt' \omega(t') \quad (58)$$

whose new coordinate satisfies

$$\partial_{t_1}^2 x_1(t_1) + \omega_1^2(t_1) x_1(t_1) = 0 \quad (59)$$

where

$$\omega_1^2(t_1(t)) = 1 + \omega^{-3/2} \partial_t^2 \omega^{-1/2}. \quad (60)$$

Clearly $\omega_1 \approx 1$ if ω varies slowly.

Renormalization consists of iterating this transformation, the aim being to freeze the frequency. The k th approximant for γ is obtained by approximating $\omega_{k+1} \approx 1$, so

$$\gamma^{(k)} = \int_{-\infty}^{\infty} (dt_{k+1} - \omega(t) dt) = \int_{-\infty}^{\infty} dt \omega \left(\prod_{j=1}^k \omega_j - 1 \right) \quad (61)$$

Thus

$$\begin{aligned} \gamma^{(0)} &= 0 && \text{(no anholonomy in this problem)} \\ \gamma^{(1)} &= \int_{-\infty}^{\infty} dt \omega(t) (\omega_1(t_1(t)) - 1) \end{aligned} \quad (62)$$

etc. $\gamma^{(1)}$ is of order ϵ .

An equivalent *Hamiltonian* renormalization is produced by iteration of the canonical transformation generated by

$$S(x, p_1, t) = x p_1 \omega^{1/2} - x^2 \partial_t \omega / 4\omega. \quad (63)$$

This gives

$$x_1 = x \omega^{1/2}; \quad p_1 = p \omega^{-1/2} + x \partial_t \omega / 2\omega^{3/2} \quad (64)$$

and hence the transformed Hamiltonian

$$\overline{H}(x_1, p_1, t) = H + \partial_t S = \frac{1}{2} \omega(t) (p_1^2 + \omega_1^2 x_1^2) \quad (65)$$

where ω_1 is given by (60). Rescaling time to t_1 as defined in (58) now gives

$$H_1(x_1, p_1, t_1) = \frac{1}{2} (p_1^2 + \omega_1^2(t_1) x_1^2(t_1)) \quad (66)$$

which is the first renormalization of the original Hamiltonian (54). The aim of subsequent renormalizations is to freeze the Hamiltonian into one whose contours are circles.

I have expressed these classical iteration schemes in terms of the renormalization of Newton's or Hamilton's equations in order to illustrate the idea behind the quantum renormalization described earlier. But they can be shown to be equivalent to the following fairly conventional WKB-like [32] procedure (to be contrasted with an unconventional WKB analysis by Wilkinson [45] which, unlike this one, does involve anholonomy). Write the exact solution of Eq. (53) as

$$x(t) = \Omega^{-1/2}(t) \cdot \exp \left\{ i \int_0^t dt' \Omega(t') \right\}. \quad (67)$$

Then the 'frequency' $\Omega(t)$ satisfies

$$\Omega^2(t) = \omega^2(t) + \Omega^{1/2}(t) \partial_t^2 \Omega^{-1/2}(t). \quad (68)$$

In terms of Ω , the phase shift is, exactly,

$$\gamma = \int_{-\infty}^{\infty} dt [\Omega(t) - \omega(t)]. \quad (69)$$

Successive approximants are obtained by the iteration

$$\Omega^{(0)} = \omega; \quad \Omega^{(k+1)} = \left[\omega^2 + (\Omega^{(k)})^{1/2} \partial_t^2 (\Omega^{(k)})^{-1/2} \right]^{1/2}. \quad (70)$$

The inevitability and universality of the divergence of these schemes can be demonstrated by considering high-order iterations of Eq. (60), for which

$$\omega_k(t) \equiv 1 + \delta_k(t) \quad (71)$$

and $\delta_k \ll 1$. Then $t_{k+1} \approx t_k$ (cf. Eq. (58)), and Eq. (60) can be written approximately as

$$\delta_{k+1}(t) \approx -\frac{1}{4} \partial_t^2 \delta_k(t). \quad (72)$$

The asymptotics of this recursion as $k \rightarrow \infty$ can be estimated by Fourier analysis, on the assumption that $\delta_0(t)$ is a real function of $\tau \equiv \epsilon t$, analytic in a strip about the real τ axis with its nearest singularities at $\tau_1 \pm i\tau_2$. Then with $\xi \equiv (\epsilon t - \tau_1)/\tau_2$ it is possible to show that

$$\delta_k(t) \xrightarrow[k \rightarrow \infty]{} \left[\frac{A \epsilon^{2k} (2k)!}{4^k \tau_2^{2k+1}} \right] \left[\frac{\cos \{ (2k+1) \cos^{-1} (1 + \xi^2)^{-1/2} \}}{(1 + \xi^2)^{k+1/2}} \right] \quad (73)$$

where A is a constant.

The first factor in (73) shows the divergence: $\epsilon^{2k} (2k)!$ decreases until $k \sim \tau_2/\epsilon$, when $\delta_k \sim \exp(-2\tau_2/\epsilon)$, and then increases until $\delta_k \sim 1$, when the scheme breaks down. The second factor is the universal function describing the asymptotic 'frequency.'

7. Historical Remarks

First I consider the important special case where the transported states $|\psi\rangle$ can be represented by wavefunctions that are *real*. Then the only possible phase factors associated with a circuit C are ± 1 . It follows [7] from the result (22) for spins that the factor is -1 when C encloses a degeneracy X^* of the spectrum to which $|\psi\rangle$ belongs; otherwise, it is $+1$. The peculiarity of this case is that parallel transport (10) is the only possible smooth continuation law, rather than a mathematically natural choice, concordant with quantum dynamics, from a infinity of possibilities.

Eigenfunctions can always be made real if their Hamiltonian matrix is real symmetric rather than complex Hermitian (this is the case when there is (bosonic) time-reversal symmetry [34]). Thus the phase law states

that an eigenfunction of a real symmetric matrix depending on parameters *changes sign* under smooth continuation round a degeneracy. This result is so simple – it holds even for 2×2 matrices – as to deserve mention in elementary expositions of matrix theory, but I have not found it in any such text. Arnold [14] is aware of the sign change, and attributes it to Uhlenbeck [35] in 1976. It was already known to theoretical chemists: Herzberg and Longuet-Higgins [36] gave an explicit statement in 1963. But the sign change (for 2×2 matrices) was implicit in work of Darboux [37] as long ago as 1896. This concerns the differential geometry of surfaces, and is worth describing.

Darboux considered a curved surface described locally by its deviation $z(X_1, X_2)$ from the plane $X = (X_1, X_2)$. Then the 2×2 real symmetric curvature matrix at X is

$$H_{ij}(X) = \partial_i \partial_j z(X). \quad (74)$$

The two eigenvalues are the principal curvatures at X , and the corresponding eigenvectors give the (orthogonal) directions of the lines of curvature at X . Degeneracies are *umbilic points*, where the surface is locally spherical (two curvatures equal). Umbilics are singularities of the net of curvature lines. The sign-change rule states that a line of curvature turns by π in a circuit of an umbilic: the Poincaré index of the tensor field (74) is $\pm \frac{1}{2}$. Fig. 4 shows how this happens for the three generic patterns [38][39] of curvature lines near an umbilic; the star has index $-\frac{1}{2}$, and the lemon and monstar have index $+\frac{1}{2}$. Star and lemon singularities occur as disclinations in liquid crystals [48].

The full phase — rather than the impoverished special case of the sign change for real matrices — was anticipated at least twice. First, in the mid-1950's, Pancharatnam [20][21][40] studied the 2-state Hermitian case in the context of the polarization states of light travelling in a fixed direction. The parameter space is the surface of the Poincaré sphere. Pancharatnam introduced the useful idea of defining two different states $|1\rangle$ and $|2\rangle$ as 'in phase' if the intensity of their superposition is maximal, a condition equivalent to their overlap $\langle 1|2\rangle$ being real and positive. This defines a connection between the corresponding parameters X_1 and X_2 as the state $|2\rangle$ obtained from $|1\rangle$ by phase-preserving transport along the shorter geodesic arc between X_1 and X_2 . He discovered that the connection is nontransitive: a circuit $X_1 X_2 X_3 X_1$ produces a state differing from $|1\rangle$ by precisely the same phase anholonomy [21] (minus half the solid angle of the circuit) as that given by parallel transport.

Second Mead [24] and Mead and Truhlar [42], studying adiabatic theory for molecules, made two important advances. They showed how the sign-change rule for degeneracies would induce modifications in the nuclear dynamics and hence change the vibration-rotation spectrum. And they realized that in the absence of time-reversal symmetry the nuclear dynamics would be influenced by the vector potential (39) and the corresponding 'magnetic' field (41), for which they gave a general formula.

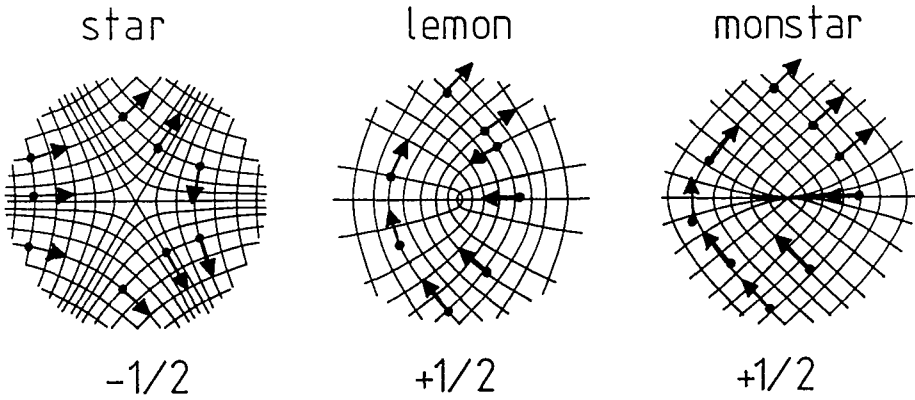


Figure 4. The generic patterns of lines of curvature near an umbilic point on a surface, illustrating the reversal ($\pm 1/2$ index) round the singularity.

My involvement with this subject began in 1979 with the appreciation [43] that degeneracies play a part in determining the fine-scale statistics of energy levels of quantum systems whose classical counterparts are nonintegrable. The systems under study possessed time-reversal symmetry and so their states should change sign round degeneracies. Seeking to display some degeneracies and their sign changes, M. Wilkinson and I [44] made a detailed investigation of the spectra of vibrating triangles as a function of angles (two parameters).

After a seminar reporting this work in the spring of 1983 at the Georgia Institute of Technology, R. Fox asked me, “what happens to the sign change if a magnetic field is switched on?”, and this question led directly to the discovery of the phase and its 2-form several weeks later. Only when the work was written in first draft was I made aware (by E.Heller) of the papers by Mead and Truhlar. In August 1983, after my paper [7] had been submitted for publication, I described the phase to B.Simon, who instantly saw its relationship to Hermitian line bundles and Chern classes. His paper [10] directed many people towards this subject, thereby provoking the considerable activity of which this book is a partial record. But thanks to a referee’s delay and an accident of astronomy, his paper appeared in 1983, mine in 1984.

Acknowledgments

I thank John Hannay for many inspirations in discussions of phase matters over several years. No military agency supported this research.

References

- [1] Landau, L.D., Lifshitz, E.M., and Pitaevskii, L.P. 1984 *Electrodynamics of Continuous Media, 2nd ed.* (vol. 8 of Course of Theoretical Physics). Oxford: Pergamon Press.
- [2] Born, M., and Wolf, E. 1956 *Principles of Optics*. London: Pergamon Press.
- [3] Chiao, R.Y. and Wu, Y.S. 1986 *Phys. Rev. Lett.* **57**, 933-936.
Tomita, A. and Chiao, R.Y. 1986 *Phys. Rev. Lett.* **57**, 937-940.
- [4] Segert, J. 1987 *Phys.Rev.* **A36**, 10-16.
- [5] Haldane, F.D.M. 1896 *Optics Letters* **11**, 730-732.
- [6] Berry, M.V. 1987 *Nature* **326**, 277-278.
- [7] Berry, M.V. 1984 *Proc.Roy.Soc.London* **A392**, 45-57.
- [8] Aharonov, Y. and Anandan, J. 1987 *Phys. Rev. Lett.* **58**, 1593-1596.
- [9] Born, M. and Fock, V. 1928 *Z.Phys.* **51**, 165-169.
- [10] Simon, B. 1983 *Phys. Rev. Lett.* **51**, 2167-2170.
- [11] Bitter, T. and Dubbers, D. 1987 *Phys. Rev. Lett.* **59**, 251-254.
- [12] Tycko, R. 1987 *Phys. Rev. Lett.* **58**, 2281-2284.
- [13] Hannay, J.H. 1985 *J.Phys.* **A18**, 221-230.
- [14] Arnold, V.I. 1978 *Mathematical Methods of Classical Mechanics*. New York: Springer.
- [15] Berry, M.V. 1986 "Adiabatic Phase Shifts for Neutrons and Photons," in *Fundamental Aspects of Quantum Theory*, eds. V.Gorini and A.Frigerio, NATO ASI series vol.144, 267-278. New York: Plenum.
- [16] Cina, J. 1986 *Chem.Phys.Lett.* **132**, 393-95.
- [17] Littlejohn, R.G. 1984 *Contemp.Math.* **28**, 151.
- [18] Berry, M.V. 1985 *J.Phys.* **A18**, 15-27.
- [19] Berry, M.V. 1983 *Semiclassical Mechanics of Regular and Irregular Motion*, in *Chaotic Behavior of Deterministic Systems*. Les Houches Lecture Series XXXVI, eds G.Iooss, R.H.G.Helleman and R.Stora. Amsterdam: North-Holland. pp.171-271.
- [20] Pancharatnam, S. 1956 *Proc.Ind.Acad.Sci.* **A44**, 247-262. Pancharatnam, S. 1975 *Collected Works of S Pancharatnam*, Oxford: University Press.

-
- [21] Berry, M.V. 1987 *J.Mod.Optics* **34**, 1401-1407.
- [22] Poston, T. and Stewart, I.N. 1978 *Catastrophe Theory and its Applications*. London: Pitman.
- [23] Berry, M.V. and Upstill, C. 1980 *Prog.Optics* **18**, 257-346.
- [24] Mead, C.A. 1980 *Chem.Phys.(Netherlands)* **49**, 23-32, 33-38.
- [25] Moody, J., Shapere, A., and Wilczek, F. 1986 *Phys.Rev.Lett.* **56**, 893-896.
- [26] Jackiw, R. 1988 *Comm.At.Mol.Phys.* **20**, 71.
- [27] Zygelman, B. 1987 *Phys.Lett.A.* **125**, 476-481.
- [28] Anandan, J. and Aharonov, Y. 1988 *Phys. Rev. Lett.*, in press.
- [29] Berry, M.V. 1987 *Proc.Roy.Soc.* **A414**, 31-46.
- [30] Garrison, J.C. 1986 Preprint UCRL-94267 from Lawrence Livermore Laboratory.
- [31] Ehrenfest, P. 1916 *Ann.d.Physik* **51**, 327-352.
- [32] Berry, M.V. and Mount, K.E. 1972 *Reps.Prog.Phys.* **35**, 315-397.
- [33] Page, D.H. 1987 *Phys.Rev.* **A36**, 3479-3481.
- [34] Porter, C.E. 1965 *Statistical Theories of Spectra: Fluctuations*. New York: Academic Press.
- [35] Uhlenbeck, K. 1976 *Am.J.Math.* **98**, 1059-1078.
- [36] Herzberg, G. and Longuet-Higgins, H.C. 1963 *Disc.Far.Soc.* **35**, 77-82.
- [37] Darboux, G. 1986 *Leçons sur la Théorie Générale des Surfaces*, vol.4, Paris: Gauthier-Villars, note VII.
- [38] Porteous, I.R. 1971 *J.Diff.Geom.* **5**, 543-564.
- [39] Berry, M.V. and Hannay, J.H. 1977 *J.Phys.* **A10**, 1809-1821.
- [40] Ramaseshan, S. and Nityananda, R. 1986 *Current Science (India)* **55**, 1225-26.
- [41] Bouchiat, C. and Gibbons, G.W. 1988 *J.Phys. France* **49**, 187-199.
- [42] Mead, C.A. and Truhlar, D.G. 1979 *J.Chem.Phys.* **70**, 2284-2296.
- [43] Berry, M. 1981 *Ann.Phys.(N.Y.)* **131**, 163-216.
- [44] Berry, M.V. and Wilkinson, M. 1984 *Proc.Roy.Soc.Lond.* **A392**, 15-43.
- [45] Wilkinson, M. 1984 *J.Phys.* **A17**, 3459-3476.
- [46] Wilczek, F. and Zee, A. 1984 *Phys. Rev. Lett.* **52**, 2111-2114.
- [47] Dingle, R.B. 1973 *Asymptotic Expansions: Their Derivation and Interpretation*, New York: Academic Press.
- [48] Frank, F.C. 1958 *Faraday Soc.Disc.* **25**, 19-28.
- [49] Bender, C.M. and Papanicolaou, N. 1988 *J.Phys.France* **49**, 561-566.
- [50] Provost, J.P. and Vallee, G. 1980 *Comm.Math.Phys.* **76**, 289-301.

THREE ELABORATIONS ON BERRY'S CONNECTION, CURVATURE AND PHASE*†

R. JACKIW

*Center for Theoretical Physics, Laboratory for Nuclear Science and Department of Physics,
Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 U.S.A.*

Received 23 October 1987

We discuss how symmetries and conservation laws are affected when Berry's phase occurs in a quantum system: symmetry transformations of coordinates have to be supplemented by gauge transformations of Berry's connection, and consequently constants of motion acquire terms beyond the familiar kinematical ones. We show how symmetries of a problem determine Berry's connection, curvature and, once a specific path is chosen, the phase as well. Moreover, higher order corrections are also fixed. We demonstrate that in some instances Berry's curvature and phase can be removed by a globally well-defined, time-dependent canonical transformation. Finally, we describe how field theoretic anomalies may be viewed as manifestations of Berry's phase.

We frequently analyze a physical system that is naturally and conveniently divided into two parts. We deal first with the motion of one set of variables, keeping the others fixed but arbitrary, and then complete the analysis of the whole by allowing variation of the previously fixed coordinates. The initially fixed variables we shall call *slow*, those whose motion is analyzed first are the *fast* variables—the terminology reflects the fact that molecular physicists and quantum chemists have used this decomposition for a long time in their Born–Oppenheimer studies of molecules. After resolving the dynamics of the fast variables, one is left with an effective action governing the slow variables. It has been established that this effective dynamics frequently involves an external vector potential \mathbf{A} which is induced by the fast variables. Moreover, the effective equation of motion satisfied by the slow variables involves only the curl of the vector potential—a magnetic-like field \mathbf{B} . Consequently, there is a gauge invariance in the description and only the gauge invariant portion of \mathbf{A} leads to physical effects. This was first seen in Born–Oppenheimer studies of molecules,¹ and analogous effects were also found in quantum field theory.² With Berry's beautiful analysis,³ we appreciate the full quantum mechanical generality of the phenomenon. The induced vector potential \mathbf{A} is now called *Berry's connection*, the induced magnetic-like field \mathbf{B} is *Berry's curvature*, while a line integral of the connection is *Berry's phase*.

* This work is supported in part by funds provided by the U.S. Department of Energy (D.O.E.) under contract number DE-AC02-76ER03069.

† Lecture delivered at "Non-Integrable Phase in Dynamical Systems," Theoretical Physics Institute, University of Minnesota, Minneapolis, Minnesota, October 1987.

Here, I shall speak on three topics: how symmetries and conservation laws are affected by Berry's connection; how higher order effects, beyond Berry's, may be determined with the help of symmetries; and finally, how Berry's connection manifests itself in quantum field theory through the anomaly phenomenon.

Let me begin by setting some notation. The complete Hamiltonian for the fast (\mathbf{p}, \mathbf{r}) and slow (\mathbf{P}, \mathbf{R}) variables is

$$H = \frac{\mathbf{P}^2}{2M} + \frac{\mathbf{p}^2}{2\mu} + V(\mathbf{R}, \mathbf{r}). \quad (1)$$

The sub-Hamiltonian governing the fast variables depends parametrically on the slow coordinates

$$h(\mathbf{R}) = \frac{\mathbf{p}^2}{2\mu} + V(\mathbf{R}, \mathbf{r}), \quad (2)$$

and possesses "instantaneous" eigenfunctions $|n; \mathbf{R}\rangle$ and eigenvalues $\varepsilon_n(\mathbf{R})$, which also depend on \mathbf{R}

$$h(\mathbf{R})|n; \mathbf{R}\rangle = \varepsilon_n(\mathbf{R})|n; \mathbf{R}\rangle. \quad (3)$$

These states can give rise to Berry's connection,³ defined by

$$\mathbf{A}_n(\mathbf{R}) = \langle n; \mathbf{R} | i\nabla_{\mathbf{R}} | n; \mathbf{R} \rangle. \quad (4)$$

In the above, I have taken the n^{th} eigenstate to be non-degenerate. Otherwise, one works within the degenerate subspace, and Berry's connection becomes a matrix in this space—an induced non-Abelian Yang-Mills-like potential.⁴

The effective Hamiltonian for slow motion, in the Born–Oppenheimer approximation, which results when off-diagonal matrix elements $\langle n; \mathbf{R} | i\nabla_{\mathbf{R}} | n'; \mathbf{R} \rangle$ are ignored, reads

$$H_{\text{eff}} = \frac{1}{2M}(\mathbf{P} - \mathbf{A}_n(\mathbf{R}))^2 + \varepsilon_n(\mathbf{R}). \quad (5)$$

Correspondingly, the effective Lagrangian is

$$L_{\text{eff}} = \frac{1}{2}M\dot{\mathbf{R}}^2 + \dot{\mathbf{R}} \cdot \mathbf{A}_n(\mathbf{R}) - \varepsilon_n(\mathbf{R}). \quad (6)$$

Equations (5) and (6) show that the fast system induces into the slow dynamics a potential energy $\varepsilon_n(\mathbf{R})$, and a velocity dependent interaction arising from Berry's connection \mathbf{A}_n , which enters dynamics only through the curvature \mathbf{B}_n associated with \mathbf{A}_n .

$$\mathbf{B}_n = \nabla \times \mathbf{A}_n - i\mathbf{A}_n \times \mathbf{A}_n. \quad (7)$$

The last term is present in the non-Abelian matrix case.

Gauge invariance of physical processes means that physical content is not affected by gauge transformations that change \mathbf{A} according to

$$\mathbf{A}(\mathbf{R}) \rightarrow g^{-1}(\mathbf{R})\mathbf{A}(\mathbf{R})g(\mathbf{R}) - ig^{-1}(\mathbf{R})\nabla g(\mathbf{R}). \quad (8)$$

[Henceforth we suppress the label “ n ” on the induced quantities.] Transformation (8) arises when the instantaneous eigenfunctions are redefined by R -dependent phase factors, which are not fixed by the instantaneous eigenvalue equation (3). Formula (8) is presented for the general case, where the connection \mathbf{A} is matrix valued, belonging to the Lie algebra of some group, and g is the matrix representation for a group element. For example, \mathbf{A} could be a Hermitian 2×2 matrix, and then g is a unitary 2×2 matrix belonging to $U(2)$. In the simplest, Abelian case, \mathbf{A} is a 1×1 matrix—a function—everything commutes and g belonging to $U(1)$ is given by $e^{i\theta}$, so that (8) reduces to the electromagnetic gauge transformation

$$\mathbf{A} \rightarrow \mathbf{A} + \nabla\theta. \quad (9)$$

The gauge transformation (8) may also be presented in infinitesimal form. Writing

$$g = I + i\Theta + \dots \quad (10)$$

where Θ is a Hermitian matrix in the Lie algebra, we deduce from (8) that the infinitesimal transformation is

$$\delta\mathbf{A} = \nabla\Theta - i[\mathbf{A}, \Theta]. \quad (11)$$

In the Abelian case, when \mathbf{A} and Θ are functions, the commutator vanishes, and (11) reduces to (9).

We are now at our first topic: Let us suppose that the full Hamiltonian (1) possesses a symmetry and consequently there exist constants of motion commuting with H . We expect that in the effective description, the symmetry is not lost and there are effective constants of motion that commute with H_{eff} . The question is how do the symmetries manifest themselves and how are the constants of motion modified in the presence of Berry's connection.

Let us for definiteness concentrate on rotational symmetry and on the associated angular momentum constant of motion \mathbf{J} . Under rotations the coordinate \mathbf{R} transforms according to $R^i \rightarrow \Lambda^{ij}R^j$ with Λ^{ij} a special orthogonal matrix. We certainly know what it means for the effective potential energy to be invariant against rotations: $\varepsilon(\mathbf{R})$ should depend only on the magnitude $R = |\mathbf{R}|$. Analytically, this requirement of rotational invariance is presented as

$$\varepsilon(\Lambda \mathbf{R}) = \varepsilon(\mathbf{R}). \quad (12)$$

For a vector quantity, like Berry's connection \mathbf{A} , the analogous requirement recognizes the vectorial nature of \mathbf{A}

$$\mathbf{A}(\Lambda \mathbf{R}) = \Lambda \mathbf{A}(\mathbf{R}). \quad (13)$$

Equivalently,

$$\Lambda \mathbf{A}(\Lambda^{-1} \mathbf{R}) = \mathbf{A}(\mathbf{R}). \quad (14)$$

Vector functions satisfying (14) have the form $\mathbf{A}(\mathbf{R}) = \mathbf{R}A(R)$, where A is a scalar function of $R = |\mathbf{R}|$. However, such connections are not physically interesting because the associated curvature (7)—the magnetic-like field—vanishes.

For a nontrivial realization of rotational invariance, we need to recognize that symmetries should be looked for in the physical content of the formalism, and we must remember that the connection has unphysical components, because a gauge transformation can always be performed without affecting physical content.

The rotational invariance requirement for a vector connection can be weakened from (14), while still retaining the full force of rotational symmetry for physical processes. Rather than demanding that $\Lambda \mathbf{A}(\Lambda^{-1} \mathbf{R})$ equal $\mathbf{A}(\mathbf{R})$, as in (14), we allow a gauge transformation to intervene.

$$\Lambda \mathbf{A}(\Lambda^{-1} \mathbf{R}) = g^{-1}(\mathbf{R}) \mathbf{A}(\mathbf{R}) g(\mathbf{R}) - ig^{-1}(\mathbf{R}) \nabla g(\mathbf{R}). \quad (15)$$

For most purposes, it is sufficient to consider the infinitesimal version of (15). When we define

$$\Lambda^{ij} = \delta^{ij} - \varepsilon^{ijk} n^k + \dots \quad (16a)$$

so that an infinitesimal rotation of \mathbf{R} is

$$\delta \mathbf{R} = \mathbf{n} \times \mathbf{R}, \quad (16b)$$

then (11) and (15) combine to

$$\mathbf{n} \times \mathbf{A} - (\mathbf{n} \times \mathbf{R} \cdot \nabla) \mathbf{A} = \nabla \Theta - i[\mathbf{A}, \Theta]. \quad (17)$$

Equation (17) may also be presented as a condition on the curvature \mathbf{B} , (7). An algebraic rearrangement of (17), which uses various vector identities, casts (17) in the form

$$(\mathbf{n} \times \mathbf{R}) \times \mathbf{B} = \nabla W - i[\mathbf{A}, W], \quad (18a)$$

$$W = \Theta + \mathbf{n} \times \mathbf{R} \cdot \mathbf{A}. \quad (18b)$$

To summarize: the effective Born–Oppenheimer Hamiltonian (5) and the effective Born–Oppenheimer Lagrangian (6) are rotationally invariant, provided, in addition to (12)—which expresses rotational symmetry of the instantaneous energy—the Berry connection \mathbf{A} satisfies (15) or (17), equivalently the Berry curvature satisfies (18)—this assures rotational symmetry up to a gauge transformation of the connection. Moreover, one can prove with general quantum mechanical reasoning that when the fast Hamiltonian $h(\mathbf{R})$ is invariant against rotations of the fast variables, supplemented by a rotation of the slow background variable \mathbf{R} , then the instantaneous eigenstates $|n; \mathbf{R}\rangle$ transform under rotations precisely in such a way that the connection defined in (4) satisfies (15).⁵ Therefore, as expected, rotational symmetry is not lost in the effective Hamiltonian, but is realized in a nontrivial fashion, when a gauge connection is present.

In the known models with nonvanishing Berry's connection, one can verify (17) or (18). Consider the example,³

$$h(\mathbf{R}) = \mathbf{S} \cdot \mathbf{R}(t) \quad (19)$$

where \mathbf{S} , playing the role of the fast variables (\mathbf{p}, \mathbf{r}), is an angular momentum operator that obeys the $SO(3)$ Lie algebra

$$[S_i, S_j] = i\epsilon_{ijk} S^k. \quad (20)$$

h is invariant against simultaneous rotations of \mathbf{S} and of \mathbf{R} , the slow background. The instantaneous eigenvalues are mR , with m ranging in unit steps from $-S$ to S . The model gives rise to Berry's curvature, which is proportional to a magnetic monopole field³

$$\mathbf{B} = -m \frac{\hat{\mathbf{R}}}{R^2}. \quad (21)$$

It is straightforward to verify that (18) is satisfied with

$$W = m\hat{\mathbf{R}} \cdot \mathbf{n}. \quad (22)$$

\mathbf{B} in Eq. (21) is an Abelian curvature. In a non-Abelian $U(2)$ example, relevant to diatoms with λ degeneracy, Berry's connection is a 2×2 matrix, which depends on a parameter κ ⁶

$$\mathbf{A}(\mathbf{R}) = \frac{1 + \kappa}{2} \frac{\hat{\mathbf{R}} \times \boldsymbol{\sigma}}{R}. \quad (23)$$

The $\boldsymbol{\sigma}$ matrices are Pauli matrices. Rotational invariance holds: \mathbf{A} satisfies (17) and Θ is given by⁷

$$\Theta = \frac{1}{2} \boldsymbol{\sigma} \cdot \mathbf{n}. \quad (24)$$

In fact one may reverse the reasoning and determine the connection or curvature from the requirements of rotational symmetry. One views (18) as an equation for \mathbf{B} and finds the most general solution in terms of W . The point is that integrability requirements on (18) sharply limit the allowed solutions. Thus the magnetic monopole in (21) is the *unique* Abelian rotationally symmetric curvature, while the configuration (23) is one of the few non-Abelian 2×2 matrix solutions. In this way knowledge of the symmetry limits the curvature, apart from the gauge freedom.

A further remark: because invariance of H_{eff} is achieved when a rotation is supplemented by a gauge transformation, constants of motion commuting with H_{eff} are modified.⁸ Recall that a constant of motion is also the generator of the infinitesimal transformation. In our case the infinitesimal rotation of coordinates must be supplemented by an infinitesimal gauge transformation of the connection. Consequently, the conserved angular momentum \mathbf{J} arising from rotational symmetry in the $\hat{\mathbf{n}}$ direction is the conventional $\mathbf{n} \cdot (\mathbf{R} \times \mathbf{P})$ supplemented by Θ , the gauge transformation generator. Equivalently,

$$\mathbf{J} = \mathbf{R} \times M\dot{\mathbf{R}} + \frac{\partial W}{\partial \mathbf{n}} \quad (25)$$

where we have used (18b) and the fact that $\mathbf{P} = M\dot{\mathbf{R}} + \mathbf{A}$. [Of course, Θ and W are linear in \mathbf{n} .] The extra term in (25) puts into evidence the physical significance of the gauge transformation that accompanies the rotational coordinate change: $\frac{\partial W}{\partial \mathbf{n}}$ is the angular momentum stored in the gauge field, that summarizes the effect of the fast variables on the slow ones. This is the reason why in the presence of a magnetic monopole [or of the Berry curvature (21)] the angular momentum contains in addition to the kinematical term the further $m\dot{\mathbf{R}}$ [see (22)]. Also this is why the $U(2)$ non-Abelian connection (23) requires that the kinematical angular momentum be supplemented by $\sigma/2$ [see (24)].

We have discussed in detail how rotational invariance and its associated conserved quantity/generator are affected by Berry's connection. However the same ideas can be applied to invariance under arbitrary coordinate transformations. When the complete Hamiltonian is invariant against some coordinate transformation, the effective Hamiltonian for slow motion retains this property, but the background gauge field induced by the fast variables is invariant only up to a gauge transformation,⁵ whose infinitesimal generator supplements the relevant constant of motion.⁸ The problem of determining all connections/curvatures that are invariant, up to a gauge transformation, against an arbitrary coordinate transformation has been solved by mathematicians and physicists.⁹

There is one more model worth mentioning: the generalized one-dimensional harmonic oscillator, with time-varying parameters

$$h = \frac{1}{2}(\alpha(t)p^2 + \beta(t)(pq + qp) + \gamma(t)q^2), \quad (26)$$

$$\alpha > 0.$$

Berry's connection is nonvanishing.¹⁰ The three operators p^2 , $pq + qp$ and q^2 close upon commutation on the Lie algebra of $SO(2, 1)$, which is like the angular momentum $SO(3)$ algebra, except some crucial signs are reversed. This may be presented by redefining the fast variables as

$$T_1 = \frac{1}{4}(q^2 - p^2), \quad T_2 = \frac{1}{4}(pq + qp), \quad T_3 = \frac{1}{4}(q^2 + p^2) \quad (27)$$

and the parameters

$$R^1 = \gamma - \alpha, \quad R^2 = 2\beta, \quad R^3 = \gamma + \alpha. \quad (28)$$

The Hamiltonian (26) takes the form

$$h = T_i R^i(t), \quad (29)$$

with

$$[T_i, T_j] = i\epsilon_{ijk} T^k. \quad (30)$$

Formulas (29) and (30) are analogous to (19) and (20) of the $SO(3)$ case, except that here a metric g_{ij} intervenes in the ijk group indices, and introduces the crucial sign differences from $SO(3)$

$$g_{ij} = \text{diag}(-1, -1, 1). \quad (31)$$

The general discussion applies, and the curvature is immediately predicted to be the unique $SO(2, 1)$ covariant,⁵

$$B^i \propto \frac{R^i}{(R^j R_j)^{3/2}} \quad (32)$$

where $R^j R_j = (R^3)^2 - (R^1)^2 - (R^2)^2 = 4(\alpha\gamma - \beta^2)$ is assumed to be positive. Equation (32) coincides with the result of the explicit calculation based on (25).¹⁰

In spite of the formal similarity to the $SO(3)$ problem, there is a crucial difference which has not been remarked upon in the literature. The curvature may be removed, not of course, by a gauge transformation, but by a globally well-defined canonical transformation, or what is equivalent by dropping a total time derivative from the Lagrangian that corresponds to the Hamiltonian (26)

$$L \equiv p\dot{q} - h = \frac{1}{2\alpha}\dot{q}^2 - \frac{1}{2}\left(\gamma - \frac{\beta^2}{\alpha} - \frac{d}{dt}\frac{\beta}{\alpha}\right)q^2 - \frac{d}{dt}\frac{1}{2}\left(\frac{\beta}{\alpha}q^2\right). \quad (33)$$

When the last term is dropped, as it can be without affecting dynamics, one is left with an equivalent theory, described by the Hamiltonian

$$\bar{h} = \frac{\alpha}{2} p^2 + \frac{1}{2} \left(\gamma - \frac{\beta^2}{\alpha} - \frac{d}{dt} \frac{\beta}{\alpha} \right) q^2, \quad (34)$$

which does not lead to a Berry connection, but is canonically equivalent to h , with the time-dependent canonical transformation $p \rightarrow p + \frac{\beta}{\alpha} q$, $q \rightarrow q$. Notice $\frac{\beta}{\alpha}$ is non-singular, because α is assumed never to vanish. Hence this canonical transformation is globally defined on the parameter space. On the other hand, an analogous transformation for the SO(3) example,³ which would rotate $\mathbf{S} \cdot \mathbf{R}(t)$ into a fixed direction, cannot be globally defined.

The reason for this difference between the SO(3) and SO(2, 1) examples derives from the fact that the parameter space [at fixed R] of the SO(3) model is the surface of a sphere, while that of the SO(2, 1) model is one sheet of the hyperboloid $R^t R_i = \text{constant}$. Unlike the former, the latter is topologically trivial, and homotopic to the Euclidean plane.¹¹

Symmetry considerations may also be used to calculate quickly and efficiently higher order corrections to Berry's phase, and this brings us to the second topic in my lecture. First we need to define what we wish to calculate in higher order. I have already indicated in Eq. (6) that the instantaneous energy eigenvalue and Berry's connection are two contributions to the effective Lagrangian induced by the fast variables onto the slow ones. Let us therefore consider the complete effective action I_{eff} induced by the fast variables. This quantity may be defined as follows. Consider the time-dependent Schrödinger equation with the time-dependent Hamiltonian $h(\mathbf{R}(t))$

$$i\partial_t |\psi; t\rangle = h(\mathbf{R}(t)) |\psi; t\rangle. \quad (35)$$

We take the two [in, out] solutions of (35) $|\psi^{(\pm)}; t\rangle$ that satisfy the initial and final conditions, respectively

$$\lim_{t \rightarrow t_i} |\psi^{(+)}; t\rangle = |n; \mathbf{R}(t_i)\rangle, \quad (36a)$$

$$\lim_{t \rightarrow t_f} |\psi^{(-)}; t\rangle = |n; \mathbf{R}(t_f)\rangle. \quad (36b)$$

For simplicity we shall assume $\mathbf{R}(t_i) = \mathbf{R}(t_f)$, but this can be relaxed. The effective action is defined from the in-out matrix element,

$$I_{\text{eff}} = -i \ln \langle \psi^{(-)}; t_f | \psi^{(+)}; t_i \rangle. \quad (37)$$

[Alternatively, I_{eff} may be given by a Feynman path integral over the fast variables.] I_{eff} is a functional of $\mathbf{R}(t)$, but it is independent of t because $|\psi^{\pm}; t\rangle$ satisfy the Schrödinger equation. Hence the overlap in (37) may be evaluated variously at $t = t_i$, where $I_{\text{eff}} = -i \ln \langle \psi^{(-)}; t_f | n; \mathbf{R}(t_i) \rangle$ or at $t = t_f$, where $I_{\text{eff}} = -i \ln \langle n; \mathbf{R}(t_f) | \psi^{(+)}; t_i \rangle$.

One may expand I_{eff} in a series of terms with increasing time derivatives of \mathbf{R} . The zeroth order term has no time derivatives; it is what would survive of I_{eff} when $\mathbf{R}(t)$ is time-independent. The first order term has a single time derivative; the second order term involves two time derivatives, etc. We now see that the instantaneous energy eigenvalue is the (negative) zeroth order term; the Berry phase gives the first order term, linear in time derivatives; higher orders are to be determined—compare (6).

$$I_{\text{eff}} = - \int dt \varepsilon(\mathbf{R}(t)) + \int dt \dot{\mathbf{R}}(t) \cdot \mathbf{A}(\mathbf{R}(t)) + \frac{1}{2} \int dt \dot{\mathbf{R}}^i(t) M^{ij}(\mathbf{R}(t)) \dot{\mathbf{R}}^j(t) + \dots \quad (38)$$

The coefficient tensors A^i and M^{ij} must be transverse to R^i . This is so because a purely radial time dependence, $\mathbf{R}(t) = \hat{R}R(t)$ with time-independent \hat{R} , gives rise only to the first term in (38). Equation (38) is an asymptotic series, it produces a real I_{eff} but misses imaginary parts, which describe decay. [Note that a second order term of the form $\int dt \dot{\mathbf{R}}(t) \cdot \mathbf{N}(\mathbf{R}(t))$ is equivalent, by an integration by parts, to the last term in (38); to justify dropping endpoint contributions, the motion must be periodic, alternatively take $t_{i,f} = \mp\infty$, where everything vanishes.]

Let me show how all higher order terms in (38) can be computed from the knowledge of the zeroth order, instantaneous eigenvalue. The discussion is confined to the SO(3) example. Observe that the generator of rotations on the fast variables \mathbf{S} , is obtained from $h(\mathbf{R})$ by differentiating it with respect to \mathbf{R} ,

$$\mathbf{S} = \frac{\partial h(\mathbf{R})}{\partial \mathbf{R}}. \quad (39)$$

This operator satisfies,

$$i[h(\mathbf{R}), \mathbf{S}] = \mathbf{R} \times \mathbf{S}, \quad (40a)$$

which is the Heisenberg picture equation for $\mathbf{S}(t)$,

$$\frac{d}{dt} \mathbf{S}(t) - \mathbf{R}(t) \times \mathbf{S}(t) = 0. \quad (40b)$$

The above non-conservation equation for $\mathbf{S}(t)$ reflects the fact that rotations on the fast variables \mathbf{S} , are not symmetry operations, unless supplemented by a rotation of the external parameters.

Next I define the in-out matrix element of \mathbf{S} ,

$$\mathbf{J}(t) = \langle \psi^{(-)}; t | \mathbf{S} | \psi^{(+)}; t \rangle / \langle \psi^{(-)}; t | \psi^{(+)}; t \rangle \quad (41)$$

which by virtue of (40a) also satisfies (40b). Finally observe that $\mathbf{J}(t)$ is given by a functional derivative with respect to $\mathbf{R}(t)$ of I_{eff} [compare with (39)],

$$\mathbf{J}(t) = -\frac{\delta I_{\text{eff}}}{\delta \mathbf{R}(t)}. \quad (42)$$

Thus combining (40) with (42) we arrive at a condition on I_{eff}

$$\frac{d}{dt} \frac{\delta I_{\text{eff}}}{\delta \mathbf{R}(t)} - \mathbf{R} \times \frac{\delta I_{\text{eff}}}{\delta \mathbf{R}(t)} = 0. \quad (43)$$

Equation (43) may be applied iteratively to (38), and then equal orders of time derivatives of $\mathbf{R}(t)$ are equated. Varying I_{eff} gives

$$\begin{aligned} \mathbf{J}(t) &= m\dot{\mathbf{R}}(t) - \dot{\mathbf{R}}(t) \times \mathbf{B}(\mathbf{R}(t)) \\ &\quad - \frac{\delta}{\delta \mathbf{R}(t)} \frac{1}{2} \int d\tau \dot{\mathbf{R}}^i(\tau) M^{ij}(\mathbf{R}(\tau)) \dot{\mathbf{R}}^j(\tau) + \dots \end{aligned} \quad (44)$$

and (43) requires

$$m \frac{d}{dt} \dot{\mathbf{R}} = -\mathbf{R} \times (\dot{\mathbf{R}} \times \mathbf{B}), \quad (45a)$$

$$\frac{d}{dt} (\dot{\mathbf{R}} \times \mathbf{B}(\mathbf{R})) = \mathbf{R} \times \frac{\delta}{\delta \mathbf{R}} \frac{1}{2} \int d\tau \dot{\mathbf{R}}^i M^{ij} \dot{\mathbf{R}}^j, \quad (45b)$$

etc. Equation (45a) determines \mathbf{B} : the known formula (21) is regained; Eq. (45b) gives a new result: the second order term in the derivative expansion, which is the first correction to Berry's phase. One finds

$$M^{ij} = \frac{m}{R^3} (\delta^{ij} - \hat{R}^i \hat{R}^j). \quad (46)$$

Clearly higher order terms may be similarly computed. The result (46) has been verified by a direct evaluation of the in, out matrix element.¹² Of course a specific value for I_{eff} is obtained only when a specific path $\mathbf{R}(t)$ is chosen and the time integral is performed.

My third and last topic concerns the role that Berry's phase plays in modern quantum field theory, where it gives another point of view on the anomaly phenomenon. This subject will be discussed in greater detail in another lecture.¹³ Here I give an introductory description, because I believe that this peculiar feature of second quantized field theories is probably unfamiliar to many of you.

I begin by reminding that the quantum mechanical revolution has not erased our reliance on the earlier classical physics. Indeed when proposing a theory, we begin with classical concepts and construct models according to the rules of classical, pre-quantum physics. We know, however, such classical reasoning is not in complete accord with quantum reality. Therefore, the classical model is reanalyzed by the rules of quantum physics, which comprise the true laws of nature, i.e., the model is *quantized*. For a long time it was believed that symmetries of a theory are not affected by the transition from classical to quantum rules. However, more recently we have learned that this is not so. In a quantized theory, some symmetries of classical physics may disappear because symmetry violating processes, which are not seen classically, can occur when the analysis is conducted with quantum effects taken into consideration. Such tenuous symmetries are said to be *anomalously* broken. Although present classically, they are absent from the quantum version of the theory, unless the model is carefully arranged to avoid this effect.

Anomalously or quantum mechanically broken symmetries play a crucial role in our present-day theories of elementary particles. In some instances they save the models from possessing too much symmetry, which would not be in accord with experiment. In other instances, the desire to preserve a symmetry in the quantum theory places strong constraints on model building and gives experimentally verified predictions. For example, the equality in the number of quarks and leptons is understood in these terms. Also, the present-day excitement about strings derives from the fact that only very few string models can be adjusted to avoid quantum mechanical, anomalous breaking of those symmetries that make string theory free of the infinities plaguing conventional field theories. Thus the number of consistent string models appears very limited, and a limitation of theoretical possibilities is what every model builder looks for. Anomalous symmetries are also beginning to play a role in other branches of physics, like condensed matter.

For a specific example of this phenomenon, consider massless fermions moving in an electromagnetic field described by electromagnetic potentials. Since massless fermions possess a well-defined helicity, we shall consider fermions with only one helicity. Such systems are an ingredient in theories of quarks and leptons. Moreover, they also arise in condensed matter physics, not because one is dealing with massless, single-helicity particles, but because a well-formulated approximation to some many-body Hamiltonians can result in a first order matrix equation which is identical to the equation for single-helicity massless fermions, i.e. a massless Dirac equation for a spinor ψ .

As a first quantized theory, the system is gauge covariant, in that a gauge transformation on the electromagnetic potential can be compensated by a change of the wavefunction, ψ . Moreover, the norm of the wavefunction is time-independent: $N = \int \psi^\dagger \psi$, $\frac{d}{dt} N = 0$. So far there are no surprises.

To construct a quantum field theory from the above, the model is second quantized: the wavefunction ψ is promoted to an operator Ψ and the state space for the theory is a many-particle Fock space. Moreover, the ground state of the second quantized

field theory has to be the filled *Dirac sea* so that the negative energy solutions of the first quantized Dirac equation are eliminated. Of course all states are functionals of the background electromagnetic potential in which the fermions move.

One expects that the second quantized theory also possesses gauge invariance, and that as a consequence of gauge invariance the total charge to which the first quantized norm is promoted, $Q = \int \Psi^\dagger \Psi$, is conserved. In fact, this is not true: the charge is not conserved; rather one finds

$$\frac{d}{dt} Q \propto \int \mathbf{E} \cdot \mathbf{H} \quad (47)$$

where \mathbf{E} and \mathbf{H} are the background electric and magnetic fields, and correspondingly gauge invariance is lost.¹⁵

There are many ways of arriving at the result where the gauge symmetry in this model is anomalously broken. In one physically transparent argument, it is established that the process of filling the negative energy sea to define the field theoretic ground state necessarily violates gauge invariance.¹⁴

Berry's ideas provide another viewpoint. The fermion field operators Ψ are thought of as fast variables, the analogs of \mathbf{p} and \mathbf{r} , or of \mathbf{S} in the example (19). The background electromagnetic potential is viewed as an external parameter; it is the analog of \mathbf{R} . The Fock-space state vectors are functionals of the background potential; they are analogs of $|n; \mathbf{R}\rangle$. An analysis shows that when the background electromagnetic potentials are gauge transformed—this can be an adiabatic change—the states acquire a phase variation and in this way lose electromagnetic gauge invariance. Thus the anomaly phenomenon is a manifestation in quantum field theory of Berry's phase, $\int d\mathbf{R} \cdot \mathbf{A}(\mathbf{R})$.¹⁶

Symmetry in quantum theory can also be seen through the realization of the symmetry algebra in the canonical commutation relations. Correspondingly, when the symmetry is anomalously or quantum mechanically broken, the algebra acquires a dynamical modification. As is seen from (6), the Berry connection induces velocity-dependent interactions, which modify the relation between canonical momentum and velocity. Consequently, the commutator of velocity components acquires a quantum mechanical correction.

$$[MR^i, MR^j] = ie^{ijk} B^k \quad (48)$$

Anomalous commutators—an important chapter in the anomaly story¹⁴—are also connected to the Berry curvature.

References

1. H. Longuet-Higgins, U. Opik, M. Pryce and R. Sack, *Proc. Roy. Soc.* **A224** (1958) 1; H. Longuet-Higgins, *Adv. Spectrosc.* **2** (1961) 429, *Proc. R. Soc.* **A344** (1975) 147; M. Child and H. Longuet-Higgins, *Phil. Trans. R. Soc.* **254** (1961) 259; G. Herzberg and H. Longuet-Higgins, *Disc. Faraday Soc.* **35** (1963) 77; M. O'Brien, *Proc. R. Soc.* **A281** (1984) 323; A. Stone, *Proc. R. Soc.* **A351** (1976) 141; C. Mead, *J. Chem. Phys.* **70** (1979) 2276, **72** (1980)

- 3839, *Chem. Phys.* **49** (1980) 23, 33; C. Mead and D. Truhlar, *J. Chem. Phys.* **70** (1979) 2284, (E)**78** (1983) 6344.
2. R. Jackiw, in *E. Fradkin Festschrift*, I. Batalin, C. Isham and G. Vilkovisky, eds. (Adam Hilger, Bristol, 1987). See also M. Asorey and D. Mitter, *Phys. Lett.* **153B** (1985) 147; Y.-S. Wu and A. Zee, *Phys. Lett.* **B258** (1985) 157.
 3. M. V. Berry, *Proc. R. Soc.* **A392** (1984) 45, and lecture delivered at this conference.
 4. F. Wilczek and A. Zee, *Phys. Rev. Lett.* **52** (1984) 2111.
 5. L. Vinet, University of Montreal preprint, August (1987).
 6. J. Moody, A. Shapere and F. Wilczek, *Phys. Rev. Lett.* **56** (1986) 893. Formula (23) is presented in a gauge which differs from the one used by Moody et al.
 7. R. Jackiw, *Phys. Rev. Lett.* **56** (1986) 2779.
 8. R. Jackiw and N. Manton, *Ann. Phys. (NY)* **127** (1980) 257.
 9. H. Wang, *Nagoya Math. J.* **13** (1958) 1; J. Harnad, S. Shnider and L. Vinet, *J. Math. Phys.* **21** (1980) 2719; R. Forgács and N. Manton, *Comm. Math. Phys.* **72** (1980) 15. For a review and more details see R. Jackiw, *Acta Phys. Austriaca*, Suppl. **XXII** (1980) 383.
 10. M. V. Berry, *J. Phys. A* **18** (1985) 15; J. Hannay, *J. Phys. A* **18** (1985) 221.
 11. P. Gerbert, MIT preprint CTP #1537, October (1987). See also E. Gozzi and W. Thacker, *Phys. Rev.* **D35** (1987) 2398.
 12. P. Gerbert, Ref. [11]; S. Iida, private communication.
 13. G. Semenoff, lecture delivered at this conference.
 14. For a review, see *Current Algebra and Anomalies*, S. Treiman, R. Jackiw, B. Zumino and E. Witten, eds. (Princeton University Press/World Scientific, Princeton NJ/Singapore, 1985).
 15. D. Gross and R. Jackiw, *Phys. Rev.* **D6** (1972) 477. For details see Ref. [14].
 16. P. Nelson and L. Alvarez-Gaumé, *Comm. Math. Phys.* **99** (1985) 103; H. Sonoda, *Phys. Lett.* **156B** (1985) 220, *Nucl. Phys.* **B266** (1986) 410; A. Niemi and G. Semenoff, *Phys. Rev. Lett.* **55** (1985) 927, **56** (1986) 1019, *Phys. Lett.* **B175** (1986) 439; A. Niemi, G. Semenoff and Y.-S. Wu, *Nucl. Phys.* **B276** (1986) 173. For reviews of this approach to anomalies, see G. Semenoff in *Super Field Theory*, H. Lee, V. Elias, G. Kunstatter, R. Mann and K. Viswanathan, eds. (Plenum, New York, 1987); A. Niemi in *Workshop on Skyrmions and Anomalies*, M. Jezabek and M. Praszalowicz, eds. (World Scientific, Singapore, 1987).