Chapter 4

Aharonov-Bohm effect and geometric phase

\[ e^{i(q/\hbar c) \oint d\mu (A_\mu + \partial_\mu \zeta)} \]

And all I wanted was a complex carrot.

"I have had my results for a long time but I do not yet know how I am to arrive at them."

Karl Friedrich Gauss
Preliminary comments

An interesting and far reaching aspect of the electrodynamics of particle-field interactions involves the electromagnetic potentials $\phi$ and $A$ and their role in the quantum mechanics of charged particles. In the last chapter, matter-radiation interactions (and corresponding spectroscopic transitions) were accounted for using the vector potential $A$. It is also important to understand how phases of quantum mechanical particle waves are affected when these waves pass through regions in which the potentials $\phi$ and $A$ are nonzero, whereas $E$ and $B$ are zero. The fields and potentials are taken to be static. The only time dependence is that arising from particle motion, and this is so slight that it can be disregarded, as explained below.

Though the Aharonov-Bohm effect is subtle, the main idea will hopefully come across. The effect relates directly to quantum electrodynamics (QED). It is germane to the gauge field theory that underlies the standard model of physics (a modest portion of which is QED), and it offers a glimpse into the weak and strong forces. For us, its significance is that it bears uncanny resemblance to the geometric phase encountered when a conical intersection of a polyatomic molecule comes into play (sometimes being encircled) through motions of the nuclei. The Aharonov-Bohm effect (hereafter referred to as the AB effect) is a good launching point for studies of conical intersections in molecules.

Like most scientific discoveries, it made its entrance amidst myriad precursor and complementary studies. It was not as original as it was in the right place at the right time. Flux quantization in superconductivity, which is similar to the magnetic version of the AB effect, had been predicted by London, refined by others, and subsumed into the finished product that was delivered by Bardeen, Cooper, and Schrieffer in 1957 (BCS theory). Ehrenberg and Siday had published an equivalent result a decade earlier (1949). The prescient 1954 paper of Yang and Mills generalized the $U(1)$ gauge symmetry of the AB effect to $SU(2)$; this paper provided the mathematical foundation of what came to be known as the Standard Model of Physics.

The 1959 paper by David Bohm and his graduate student Yakir Aharonov is about quantum mechanical effects that arise when particles pass through regions where the

‡ David Bohm (1917-1992) was a major figure in 20th century physics. He also led an extraordinary life: maligned during the McCarthy witch-hunt era (including being arrested and being fired by Princeton University) for his 1930s participation in political organizations such as the Young Communist League; moving from the U.S. to Brazil, then Israel, then England; contributing to the fledgling field of neuropsychology, including development of the holonomic theory of the brain; toward the end of his life suffering from severe depression. In 1959, he and his student Yakir Aharonov published their paper. Shortly thereafter, they learned that Ehrenberg and Siday had derived the same result a decade earlier. Consequently Bohm referred to the ESAB effect. This did not stick and the effect carries the names Aharonov and Bohm.
Aharonov-Bohm effect and geometric phase

potentials $\phi$ and $A$ are nonzero, whereas the so-called physical fields $E$ and $B$ are zero. It raised considerable interest and speculation, with debates that raged for years, some continuing to this day. I cannot explain why it created such a controversy, except to point out that such things happen from time to time in science.

The math used in the previous chapter was appropriate for 3D objects (vectors, fields, etc.), with time handled separately from spatial displacement. This non-covariant approach will continue, as it accommodates the largest audience. Occasions arise, however, when it is prudent to introduce generalizations, complementary perspectives, and extensions that enlist covariant spacetime and more sophisticated concepts. The more challenging parts will be indicated as such.

1. The main issue

In classical physics, a particle that moves in vacuum in the presence of an electromagnetic field experiences a force described by the Lorentz force equation:

$$F = q\left( E + (\nu/c) \times B \right)$$

where $q$ is the charge of the particle, and $\nu$ is its velocity.

The potentials $\phi$ and $A$ are introduced as a means of obtaining $E$ and $B$. They have no physical significance of their own. Indeed, they serve only as convenient means of obtaining the so-called physical fields $E$ and $B$, which are solely responsible for the electrodynamics forces. In fact, the potentials are not even unique. The fields $E$ and $B$ are obtained by differentiating them, so they can be altered without affecting $E$ and $B$. These alterations are the gauge transformations discussed earlier. For example, with $B = \nabla \times A$, adding the gradient of a scalar to $A$ does not affect $B$. Forces other than the electrodynamics force given by eqn (1.1) might be present as well. Their inclusion yields the equations of motion.

Alternatively, a Hamiltonian can be used to obtain the equations of motion. It is invariably expressed in terms of $\phi$ and $A$. The equations of motion are obtained by using Hamilton's equations, and the resulting coupled equations contain derivatives of $\phi$ and $A$. However, a bit of skilled manipulation converts these equations to ones in terms of $E$ and $B$. The Lorentz force given by eqn (1.1) is recovered, illustrating the consistency of the two approaches. Nothing could be more straightforward, at least from a classical perspective.

Quantum versus classical

Quantum mechanics, on the other hand, uses wave functions to describe particle waves and the dynamical processes they undergo. Sometimes the explicit use of wave functions can be avoided (or is not relevant, as with spin), but wave functions nonetheless lie at the heart of the theory. Phase is present from the outset: from the phase factor $\exp(-iE_n t/\hbar)$ of the $n$th eigenstate, to the relative phases that enter the construction of a

---

5 Speaking of uniqueness, note that $E$ and $B$ have a total of 6 components, whereas $\phi$ and $A$ have a total of only 4 components.
wave function is zero) the particle experiences the potential to enter. Instead of zero everywhere, it is brought about by the voltages where they are set at fixed values, say \( \phi_1^0 \) and \( \phi_2^0 \). Before the packets leave the cylinders, the voltages are turned back to zero. The upper and lower packets each acquire a phase that is brought about by the voltages \( \phi_1^0 \) and \( \phi_2^0 \). These phases are \( e^{i\phi_i^0 \tau / \hbar} \) and \( e^{i\phi_2^0 \tau / \hbar} \), where \( \tau \) is the time that the voltage is on. Clearly the phases differ if \( \phi_1^0 \neq \phi_2^0 \).

![Figure 1](image1.png)

Figure 1. (a) An incident wave packet is split into two coherent wave packets that pass through perfectly conducting cylinders, i.e., regions where \( E \) is always zero. The electric potentials \( \phi_1 \) and \( \phi_2 \) are slowly turned on after the packets have entered the cylinders, reaching steady values of \( \phi_1^0 \) and \( \phi_2^0 \). They are slowly turned off before the packets exit the cylinders. Each of the packets acquires phase, and they are made to recombine at the screen labeled interference. The interference pattern changes as the difference between the potentials is varied. (b) Waves pass through regions where \( B \neq 0 \), whereas \( A_1(r) \neq 0 \) and \( A_2(r) \neq 0 \). Of course there exists some region where \( B = 0 \), but the waves do not access this region.

The electric field, though it is always zero where the wave function is present, is not zero everywhere. It is nonzero in a region of space that the wave function is not allowed to enter. Instead of experiencing \( E \) directly (i.e., the local \( E \) field experienced by the wave function is zero) the particle experiences the potential \( \phi \) associated with \( E \).
The situation shown in Fig. 1(b) is similar in spirit. Waves that follow the upper and lower paths acquire phases as they pass through the $A_1(r) \neq 0$ and $A_2(r) \neq 0$ regions, as discussed below. The sketch in Fig. 1(b) is ambiguous, however, because it says nothing about the location of the $B \neq 0$ region. We will see that if it lies inside the region enclosed by the arrows interference is affected by the strength of $B$, whereas if it lies outside this region, varying $B$ has no affect on the interference pattern.

The electric version of the AB effect is not as easily verified experimentally as the magnetic version. Perhaps for this reason, the effect that derives from the presence of the vector potential $A$ is usually referred to as the AB effect. It is worth noting that there must be both magnetic and electric versions. In special relativity, electric and magnetic fields are transformed into each other via Lorentz transformation. Though it is interesting to ponder the electric and magnetic versions (separately or together), here we shall consider just the magnetic one.

It is noteworthy that quantum mechanical waves respond to the potentials, which are not gauge invariant. For example, if $A$ is changed by the addition of $\nabla \zeta$, where $\zeta$ is a scalar function, the quantum mechanical wave must change accordingly in order that the gauge transformation: $A \rightarrow A + \nabla \zeta$, has no physical consequence. This feature, in which $A$ appears in a phase along a trajectory, distinguishes classical and quantum particles. As you might imagine, everything works out properly.

2. $E$ and $B$ versus $\phi$ and $A$

When a particle wave packet passes through a region where $E = 0$ and $B = 0$, the Lorentz force is zero, so there cannot be an effect that has a classical counterpart. Quantum mechanically, however, relative phases of wave packets can be manifest in interference phenomena that have no classical counterparts. Consequently, we need to look carefully at the roles played by potentials, for example, in the context of the Schrödinger equation.

---

It might seem that the effect on the quantum system is non-local. After all, something happens because of $B$ that exists elsewhere but is zero in the region of the quantum system. Though the Lorentz force being zero might entice one to think that nothing can happen, such instinct is based on classical physics, which does not take into consideration interference of particle waves. The physics is indeed local: it is brought about through the gauge field $A$.

Another example of a debate over local versus non-local physics is the Einstein-Podolsky-Rosen paradox. Consider two atoms that are created by photodissociation of a homonuclear diatomic molecule. The atoms move in opposite directions in vacuum until they are far apart. Then one of them is detected. It is known that one atom is formed in state $a$, while the other is formed in state $b$. The state $\psi$ that describes the system gives equal probability of finding the detected particle in either of the states $a$ or $b$: $\psi \sim \psi_1(a) \psi_2(b) \pm \psi_1(b) \psi_2(a)$. The point is this. If particle 2 is detected in state $a$, particle 1 is instantly collapsed into state $b$. Information is not transferred to particle 1 at the speed of light; the collapse is instantaneous. In other words, the state of particle 1 is determined by an act that takes place far away from particle 1. This has been presented as a non-local effect. However, it is not mysterious, nor is it non-local. The state of the system must include the relative translational motion that connects the two atoms. The state of the system happens to be physically enormous, as it contains a function for fragment relative translational motion.

213
Aharonov-Bohm effect and geometric phase

The debate about which pair is more fundamental: $E/B$ or $\phi/A$, will be avoided. The notion that one must choose between these options is what spawned and kept alive the original controversy over the AB effect. Before the 1959 paper, it had been accepted by the majority of scientists that potentials are conveniences that have no physical significance of their own. However, the AB effect showed that measurable effects could be attributed to the potentials. Consequently, it was suggested that privileged status should be assigned to $\phi$ and $A$ rather than to $E$ and $B$. Controversy ensued. No surprise: throughout the tortured evolution of electrodynamics even great scientists got things wrong from time to time.

The question itself misses the mark. In the classical theory, the vector potential is optional, a convenience so to speak. On the other hand, in quantum mechanics the important equations contain $\phi$ and $A$, not $E$ and $B$. This is no accident — the potentials are essential.

We shall now consider the effect of $A$ on a charged particle wave function. It is assumed that the particle does not interact with other particles, and that it experiences no potential other than $A$, i.e., $\phi$ is zero. To distinguish effects due to $B$ from effects due to $A$, an experimental arrangement is conjured in which the region of space under consideration has $B = \nabla \times A = 0$, whereas $A \neq 0$.

The phase of an electron wave function is affected as it passes through the $A \neq 0$ region. However, if this is all that happens (i.e., there is no boundary condition to be satisfied and the wave function only acquires a phase shift) there will be no observable effect because $|\psi|$ is unaffected. On the other hand, interference depends on relative phase. Thus, observable effects can arise in cases in which two waves that are coherent with respect to one another pass through a region of nonzero $A$ and are then superposed. For example, in Fig. 1(b) each wave in general experiences a different $A$. Following passage through their respective $A \neq 0$ regions these waves are brought together such that interference is observed. In other words, $|\psi_1 + \psi_2|$ is affected by the relative phase between $\psi_1$ and $\psi_2$. Because such particle interference has no classical counterpart, effects arise due to the presence of $A$ in regions where $B = 0$. This is the basis of the AB effect.

3. Particle-on-a-ring

Let’s now examine a charged particle confined to a circular ring in a region where $B = 0$ and $A \neq 0$. The experimental arrangement suggested in the 1959 paper by Aharonov and Bohm will be analyzed later. The presence of nonzero $A$ requires that there exists some region of space where $B \neq 0$. However, the particle wave is forbidden to enter this region.

It is shown in Appendix 3 that a charged particle is affected by the presence of $A$. For example, this is seen with the kinetic energy operator:

$$T = (p - (q/c)A) \cdot (p - (q/c)A) / 2m$$  \hspace{1cm} (3.1)

where $p = -i\hbar \nabla$ is the canonical momentum. This is standard quantum mechanics, albeit in the present context with a classical rather than quantized electromagnetic field.
Aharonov-Bohm effect and geometric phase

In the particle-on-a-ring (POR) example discussed below, it is seen that $A$ lifts the twofold degeneracy of the field-free $e^{\pm i\phi}$ pairs. In other words, when $A \neq 0$ the energies are not the same for $e^{i\phi}$ and $e^{-i\phi}$. For an observable effect to exist, it is essential that the region of nonzero $B$ is inside the red ring shown in Fig. 2. Were the region of nonzero $B$ outside the ring, there would be no observable effect, despite the fact that $A$ is nonzero on the ring. The reason for this will be made clear in subsequent sections. Following this exercise, a succinct way to account for $A$’s presence is introduced. Namely, in regions where $B = 0$, to obtain the effect brought about by $A$, we need only calculate a phase by integrating $A$ over a convenient path.

These exercises enable the topological difference between simply connected and non-simply-connected domains to be introduced in a transparent manner. It is hoped that they enable the AB effect to be seen as unambiguous, even intuitive.

As mentioned earlier, an important application of the results obtained in this chapter arises in the area of molecular electronic structure and associated nuclear dynamics. It is known that conical intersections of potential energy surfaces abound in nature and have important consequences: radiationless decay (internal conversion, intersystem crossing), chemical reaction pathways, etc. Something important called geometric phase accrues when a conical intersection is approached (sometimes encircled) through motions of the nuclei. For example, geometric phase can affect vibrational wave functions profoundly. Without its inclusion they are often egregiously in error, even qualitatively.

The phase associated with the AB effect is, for all practical purposes, the same as this molecular geometric phase. Therefore, as a matter of pedagogy, it makes no sense to approach conical intersections in electronic structure theory without first taking the time to understand the AB effect. In examining both the AB effect and conical intersections, we will see that intuition flows both ways, as the latter can provide insight into the former.

The chapter ends with a summary that, among other things, puts to rest the paradox: If $E$ and $B$ are the true physical fields and $\phi$ and $A$ are mere constructs that serve only to yield $E$ and $B$, how can $\phi$ and $A$ be responsible for an experimental effect? We will see that this line of thought is inconsistent with quantum mechanics.

**4. Solution for $A \neq 0$**

Referring to Fig. 2, assume that a magnetic field $B = B_r^2$ of uniform strength is contained within a cylinder (solenoid) of radius $a$. A charged particle moves on a ring of radius $b$. Outside the cylinder $B = 0$, so the total magnetic flux $\Phi$ is
\[ \Phi = \oint_{S_c} dS \cdot B = \pi a^2 B = \oint_{S_c} dS \cdot \nabla \times A \] (4.1)

Application of Stokes' theorem yields the relationship between \( \Phi \) and the closed line integral of \( A \) around the ring: \(^1\)

\[ \Phi = \oint_C d\ell \cdot A = b \oint_C d\phi A_\phi = 2\pi b A_\phi \quad \Rightarrow \quad A_\phi = \Phi / 2\pi b \] (4.2)

The time independent Schrödinger equation, using \( A_\phi = \Phi / 2\pi b \) from eqn (4.2), and with canonical momentum: \( p = p_\phi \hat{\phi} = -i\hbar (\partial / \partial \phi) \hat{\phi} \) is

\[ E\psi = \frac{1}{2m} \left( p + \left( e/c \right) A \right)^2 \psi = \frac{1}{2m} \left\{ \frac{\hbar^2}{b^2} \frac{\partial^2}{\partial \phi^2} + \left( \frac{e\Phi}{2\pi bc} \right)^2 - i\hbar \left( \frac{e\Phi}{\pi bc} \frac{\partial}{\partial \phi} \right) \right\} \psi \] (4.3)

The charge in eqn (4.3) has been taken to be that of an electron, \( i.e., \quad q = -e \) (where \( e = +1.6 \times 10^{-19} \) Couloms, SI units). The fact that \( p \cdot (A\psi) \) is equal to \( A \cdot (p\psi) \) (\( i.e., \ A = A_\phi \hat{\phi} \) is constant) is used to write \( p \cdot A + A \cdot p = 2A \cdot p \), which in the present case is equal to \( 2A_\phi p_\phi = 2(\Phi / 2\pi b)(-i\hbar \partial / \partial \phi) \). When this is multiplied by \( e/c \), the result is recognized as the rightmost term inside the bracket in eqn (4.3). Equation (4.3) is tidied by introducing constants: \( hc/e = \Phi_L \) is the London flux quantum (which arises in superconductivity), \(^5\) and \( \hbar^2 / 2mb^2 = B_{rot} \) is the familiar rotational constant. Thus, eqn (4.3) becomes

\[ \frac{d^2 \psi}{d\phi^2} + i2(\Phi / \Phi_L) \frac{d\psi}{d\phi} + C\psi = 0 \] (4.4)

where \( C = E / B_{rot} - (\Phi / \Phi_L)^2 \). The solution is obtained by introducing \( \psi \propto e^{in \phi} \), which yields: \( n^2 + 2(\Phi / \Phi_L)n - C = 0 \). Thus,

\[ n = -\Phi / \Phi_L \pm \sqrt{E_n / B_{rot}} \] (4.5)

To satisfy the boundary condition: \( \psi(0) = \psi(2\pi) \), \( n \) must be an integer. Consequently, the energy eigenvalues are (see Fig. 3)

\[ E_n = B_{rot} \left( n + \Phi / \Phi_L \right)^2 \] (4.6)

\(^1\) Symmetry ensures that \( A \) has no \( \phi \) or \( z \) dependence; only radial dependence is allowed. Because \( \nabla \times A = 0 \) outside the cylinder, \( A_z \) is constant for \( r > a \) so it can be set to zero. Likewise, \( A_r \) is constant on the ring so it can be set to zero.

\(^5\) In superconductivity the appropriate quantum is \( \Phi_0 = \Phi_L / 2 \). The London quantum \( \Phi_L \) was introduced before it was appreciated that current is carried by electron pairs. In a BCS superconductor, charge is carried by a quasiparticle called a Cooper pair. Its charge is \( -2e \) instead of the electron charge \( -e \) of a regular conductor.
Aharonov-Bohm effect and geometric phase

For each value of \(n\), the wave function (leaving aside normalization) is

\[
\psi_n = e^{i n \phi} \tag{4.7}
\]

Note that energies for \(n\) values that have the same magnitude but differ in sign are not degenerate, as they are when \(A = 0\). For example, putting \(n = +2\) and \(-2\) alternately into eqn (4.6) yields

\[
E_{+2} = B_{rot} \left( 2 + \frac{\Phi}{\Phi_L} \right)^2 \\
E_{-2} = B_{rot} \left( -2 + \frac{\Phi}{\Phi_L} \right)^2 \tag{4.8}
\]

This difference is seen in Fig. 3 for the case \(\Phi/\Phi_L = 1\); clearly \(E_{+2} \to E_{-2}\) as \(\Phi \to 0\).

The canonical and kinetic angular momenta differ. The former is obtained by differentiating eqn (4.7), which yields \(n\hbar\). The latter includes the vector potential and is given by \(\hbar(n + \Phi/\Phi_L)\). Clockwise and counterclockwise directions have different speeds for a given value of \(|n|\). An interesting feature can be seen with eqns (4.6) and (4.8). Namely, energies are repeated if \(\Phi/\Phi_L\) is an integer. For example, with \(\Phi/\Phi_L = 1\), \(E_{+2} = 9B_{rot} = E_{-4}\). In the present context this appears as a curiosity, but in superconductivity it is important. If a large superconducting current flows on the ring it will produce a large magnetic field inside the ring. In other words, there need be no external source of magnetic field. The resulting flux \(\Phi\) is quantized. This is sufficiently interesting (and closely related to the AB effect) that it is presented and discussed in Appendix 5: Flux quantization in superconductivity.

The vector field \(A\) causes the electron wave's phase to change as the wave circulates on the ring. To satisfy the boundary condition, this phase change in turn results in higher or lower kinetic momentum.

The above example illustrates how a phase that has no classical counterpart has a significant effect on a quantum mechanical system.

5. Vector potential appears in a phase factor

Let us now return to the time dependent Schrödinger equation for a particle of charge \(q\) in a region where \(A \neq 0\) and \(B = 0\).

\[
\left( \frac{1}{2m}(-i\hbar \nabla - (q/c)A)^2 \right) \psi = i\hbar \frac{\partial \psi}{\partial t} \tag{5.1}
\]
The scalar potential has been set to zero. The solution to eqn (5.1) is given by the \( A = 0 \) solution times a phase factor. Specifically,

\[
\psi = \psi_0 e^{ig(r)}
\]

where \( \psi_0 \) is the \( A = 0 \) solution. To verify that the wave function given by eqn (5.2) indeed is the solution to eqn (5.1), take the gradient of \( \psi \):

\[
\nabla \left( \psi_0 e^{ig(r)} \right) = \left( \nabla \psi_0 \right) e^{ig(r)} + i \frac{\nabla g(r)}{q/\hbar c} \psi_0 e^{ig(r)}
\]

This result does not depend on the path taken between \( r_0 \) and \( r \), just the integration end points. This is true in any region where \( \nabla \times A = 0 \), as can be seen by applying Stokes’ theorem. The underlying reason for the path independence is discussed in the next section. Using eqn (5.3), the kinetic momentum \( \pi \) operating on \( \psi = \psi_0 e^{ig(r)} \) becomes

\[
\pi \psi = \left( -i\hbar \nabla - (q/c)A \right) \psi_0 e^{ig(r)}
\]

\[
= \left\{ -i\hbar \left( \nabla \psi_0 \right) - ih \left( i(q/c)A \right) \psi_0 - (q/c)A \psi_0 \right\} e^{ig(r)}
\]

\[
= -i\hbar \left( \nabla \psi_0 \right) e^{ig(r)}
\]

Note: two terms that each contain \( A \) have cancelled one another. To obtain \( \pi \cdot \pi \psi \), operate from the left on eqn (5.4) with \( \pi = \left( -i\hbar \nabla - (q/c)A \right) \). A little algebra yields

\[
\left( -i\hbar \nabla - (q/c)A \right)^2 \psi = -\hbar^2 \left( \nabla^2 \psi_0 \right) e^{ig(r)}
\]

When this is put into eqn (5.1) the exponentials \( e^{ig(r)} \) cancel, leaving a Schrödinger equation for \( \psi_0 \):

\[\text{Aharonov-Bohm effect and geometric phase}\]

\[\text{The introduction of this form is not as ad hoc as it might appear. Phase accumulated along a path is given by the action, } S, \text{ which is the integral of } dt L:\]

\[S = \int dt L = \int dt \left( T + (q/c)\mathbf{v} \cdot \mathbf{A} \right) = S_0 + \frac{q}{c} \int \mathbf{r}' \cdot \mathbf{A}(\mathbf{r}')\]

The subtlety is the Lagrangian: \( T + (q/c)\mathbf{v} \cdot \mathbf{A} \), which can be found in Appendix 3, eqn (20). Beyond elementary systems, Lagrangians are often conjured to yield the equations of motion. It is straightforward (but algebraically tedious) to verify that this Lagrangian yields the correct equations of motion.

\[\text{\dag \dag \text{The Leibniz integral rule for differentiation of an integral is:}}\]

\[\frac{d}{d\alpha} \int_{a(\alpha)}^{b(\alpha)} dx f(x, \alpha) = \frac{db(\alpha)}{d\alpha} f(b(\alpha), \alpha) - \frac{da(\alpha)}{d\alpha} f(a(\alpha), \alpha) + \int_{a(\alpha)}^{b(\alpha)} dx \frac{\partial}{\partial \alpha} f(x, \alpha).\]
Aharonov-Bohm effect and geometric phase

\[-\frac{\hbar^2}{2m} \nabla^2 \psi_0 = i\hbar \frac{\partial \psi_0}{\partial t}\] (5.6)

Equation (5.6) confirms that \(\psi_0\) is the solution to a Schrödinger equation obtained by setting \(A = 0\). These manipulations verify that the effect on the wave function brought about by \(A \neq 0\) in a region where \(B = 0\) is accounted for with a phase factor. As mentioned earlier and discussed below, this result is valid only in regions where \(B = \nabla \times A = 0\).

6. Topology

Referring to Fig. 4, Stokes' theorem is applied to a region where \(\nabla \times A = 0\). Because all closed line integrals of \(A\) are equal to zero in this region, the line integral between points \(A\) and \(B\) depends only on the end points, not the path between them. Were \(\nabla \times A\) not zero, circuit integrals would enclose flux, and integration along different paths would in general yield different results. In this case, eqns (5.2) and (5.4) would not apply. This is why the material in the last section applies only to regions where \(\nabla \times A = 0\).

Because of the fact that only the end points count, the paths can be distorted to suit our convenience. For example, a complicated trajectory that passes around a solenoid can be replaced with a simpler one. This is the situation that exists in a simply connected region. A simply connected region is one in which a path that surrounds the region (in this case a path for a line integral) can be continuously deformed until it becomes a point.

As mentioned above, in a region where \(\nabla \times A \neq 0\), integration is over a surface that flux passes through, so Stokes' theorem yields \(\Phi_{\text{enc}} = \frac{1}{\hbar} \oint dr \cdot A(r)\). Thus, a line integral between two points is not independent of the path, because different closed circuits enclose different amounts of flux. For example, when the path surrounds a cylinder that contains flux, we get the results that were obtained in the POR section. In this case, the closed circuit integral encloses a non-simply-connected region.

7. Aharonov-Bohm

The groundwork has been laid. An experimental geometry close to that suggested in the original paper is shown in Fig. 5. A particle wave incident from the left is split into two equally intense components. The \(B\) field (out of the page) is present in a cylindrical solenoid. It is assumed that \(B\) is zero outside the cylinder. One component wave passes above the cylinder, while the other passes below it. Outside the cylinder, \(A\) is nonzero, so the waves experience \(A\), but not \(B\). It is assumed that passage is sufficiently slow that the fields are perceived by the particle as static, \textit{i.e.}, no transitions between energy levels
arise because of a time varying electromagnetic field. Of course a free electron has no energy levels other than spin, which is irrelevant in the present context. However, had we transported an electron in a box (vide infra, Section 9), the issue of transitions induced by a time varying electromagnetic field would have arisen.

![Diagram of electron interference](image)

Figure 5. The $B$ field (out of the page) is fully contained within a cylinder, whereas $A$ is nonzero outside the cylinder. Waves that pass above and below are obtained by splitting a single wave incident from the left in order to ensure coherence. That is, with $B = 0$, they retain a fixed phase with respect to one another throughout the trajectory. With $B \neq 0$, a phase shift between them is acquired that depends on the strength of $B$ and is detectable via interference at the place where the beams are recombined. As the strength of $B$ is varied, the interference pattern changes. Interestingly, the "center-of-mass" of the interference pattern is unaffected, as moving this center-of-mass would require a force.

From the electron's perspective it is inevitable that $A$ is, to some extent, time dependent because the electron enters and leaves the region of nonzero $A$. This causes the electron to experience an electric field ($-\partial A/\partial t$) and, from Maxwell's equations, a magnetic field. However, this has a negligible effect. Several authors have shown that the $E$ and $B$ fields thus generated are inconsequential. Verification on your part is left as an exercise.

As shown above, in the region $r > a$, where $a$ is the cylinder radius, $A$ is given by

$$A = \frac{\Phi}{2\pi r} \hat{\phi}$$

(7.1)

where $\hat{\phi}$ points counterclockwise.

At the points of closest approach to the cylinder, for the lower component wave, $A$ points in the same direction as the wave's momentum, while for the upper component wave $A$ is directed oppositely to the wave's momentum. Phase shifts are introduced to each of the waves, in one case advancing the wave's phase, in the other case retarding it.

After passing through the region near the cylinder, the waves are guided such that they recombine. Assuming that coherence has been maintained throughout the transit from $r_1$ to $r_2$, the interference pattern depends on the flux $\Phi$. Thus, a reasonable experiment is to detect signals due to the recombining upper and lower waves and monitor them as $B$ is varied. This interference region is indicated by the thin black rectangle in Fig. 5. For example, this will result in a sinusoidal modulation of the signal at a given location. Movies that show how this works are at http://rugth30.phys.rug.nl/quantummechanics/ab.htm.
The presence of $A$ results in a phase difference, $\beta$, for waves that follow the two paths indicated in Fig. 5. Using eqn (5.2) yields

$$\beta = \frac{q}{\hbar c} \left( \int_{r_1}^{r_2} d\mathbf{r} \cdot A_{\text{lower}} - \int_{r_1}^{r_2} d\mathbf{r} \cdot A_{\text{upper}} \right) \quad (7.2)$$

In reversing the limits on the second integral in order to change its sign, we see that $\beta$ is the same as for a single wave that traverses the following circuit: (i) start at $r_1$; (ii) pass below the cylinder; (iii) turn around at $r_2$; (iv) pass above the cylinder; and (v) arrive back at $r_1$. Thus,

$$\beta = \frac{q}{\hbar c} \oint d\mathbf{r} \cdot A \quad (7.3)$$

The right hand side of eqn (7.3) is the closed circuit integral of a momentum: (i.e., an action) divided by $\hbar$. Because $\nabla \times A = 0$ outside the cylinder, paths between $r_1$ and $r_2$ can be deformed as we please. For example, deformations that capitalize on the symmetry are obtained by using semicircular paths of radius $a + \epsilon$, where the solenoid radius is $a$ and $\epsilon$ is small: one above the cylinder, the other below (Fig. 6). Integration over the closed circuit yields

$$\beta = \frac{q}{\hbar c} \Phi = 2\pi \Phi / \Phi_L \quad (7.4)$$

This result is the same as the one obtained using the POR model.

The integration in eqn (7.3) is written for a closed circuit. Keep in mind, however, that there is no boundary condition to be met, in contrast to the POR case. The circuit $C$ is complete in the sense that the mathematics can be written as a closed line integral, but there is no bound eigenstate. Thus, $A$ does not affect the energy of the waves that pass above and below the cylinder. A phase shift arises because $A$ appends phases to the waves, and this is manifest in the interference indicated in Fig. 5.
8. Extensions and generalizations

The potentials \( \phi \) and \( A \) have emerged as the players in the quantum theory. There is no temptation to use \( E \) and \( B \), nor would they suffice were this attempted. Without doubt, the quantum mechanical wave of a particle having charge \( q \) acquires phase when it passes through a region of nonzero \( \phi \) and/or \( A \):

\[
\exp\left(-i\frac{q}{\hbar}\int dt \phi\right)
\]

(8.1)

and

\[
\exp\left(i\frac{q}{\hbar c}\int d\mathbf{r} \cdot \mathbf{A}\right)
\]

(8.2)

It is no coincidence that the expressions given by eqns (8.1) and (8.2) can be combined into a single expression by using the following 4-vector:

\[
A^\mu = (A^0, A^1, A^2, A^3) = (\phi, \mathbf{A})
\]

\[
= (\phi, A_x, A_y, A_z) = (A_0, -A_1, -A_2, -A_3)
\]

(8.3)

We will continue using boldface type to represent 3-vectors, while 4-vectors will be represented using a normal font and Greek superscripts and subscripts where appropriate. Until now, we have referred to the \( E/B \) pair as fields and the \( \phi/A \) pair as potentials. Of course, the potentials are also fields according to the mathematical definition of a field. Now that we have, for the most part, abandoned \( E/B \) in favor of \( \phi/A \), the potentials will be referred to as fields. For example, the gauge field of electromagnetism is \( A^\mu \). Note: In dealing with phase and (covariant/contravariant) tensor algebra, it is hard to avoid sign errors. Do not be disheartened if this becomes exasperating — it is just math.

The product \( dx^\mu A_\mu = dx_\mu A^\mu \) is a Lorentz scalar. Thus, the central object in the theory is invariant with respect to Lorentz transformation. The other important requirement is gauge invariance. Physical quantities cannot depend on a choice of gauge. A schematic illustration of the gauge field and how gauge invariance works is given in Fig. 7. The plane denoted \( x^\mu \) is a schematic representation of 4D spacetime. From each spacetime point a strand called a fiber extends upward. In Fig. 7(a), the blue dots represent the gauge field associated with spacetime points. The gauge field in general varies in spacetime, so the points are shown at different heights. The spacetime points are of course densely packed, as they must be if they are to represent a continuous spacetime. The collection of fibers is called a bundle. The mathematical term fiber bundle brings to mind the bundles of fibers that are used in fiber-optic communications.
The exercises carried out in the last few sections have shown that the gauge field appears in both the kinetic momentum operator and the state vector. This is essential if gauge invariance is to be satisfied. Referring to Fig. 7(b), a local gauge transformation applied to the gauge field is represented schematically by displacements of the blue dots along their respective fibers. At the same time, the wave function must change in a correlated manner (by acquiring a phase factor) to ensure that the physical world remains oblivious to our choice of gauge. This phase factor is indicated with red dots and yellow circles, i.e., phase angles are defined relative to the axes shown. Though Fig. 7 is highly schematic, it gives one a sense of the registry that must exist between the gauge transformation of the field \( A^\mu \) and the gauge transformation of the wave function.

To see how this works insofar as wave functions and observables are concerned, let us add to \( A \) the gradient of a scalar: \( A \rightarrow A + \nabla \zeta \). We know that this does not affect \( B \), because \( B = \nabla \times A \) and the curl of a gradient is identically zero. Also, \( \partial(\nabla \zeta)/\partial ct \) must be added to \( E \) to cancel the contribution that arises from changing \( A \). In other words, \( \hat{E} \) remains \(-\nabla \phi - \partial A / \partial ct\), where \( \phi \) and \( A \) are the original potentials, because the gauge transformation gives: \(-\nabla \phi - \partial(\hat{A} + \nabla \zeta)/\partial ct + \partial(\nabla \zeta)/\partial ct\), and the \( \nabla \zeta \) terms cancel, leaving the original \( \hat{E} \).

The effect on \( \psi \) of adding \( \nabla \zeta \) to \( A \) is given by

\[
\psi \rightarrow \tilde{\psi} = \psi e^{i(q/\hbar)c} \zeta
\]  

(8.4)

When the gauge transformed kinetic momentum operator \( \hat{\pi} \) acts on \( \tilde{\psi} \) given by eqn (8.4) we find that nothing has changed. In other words, the gauge transformation does not affect the physical world. We would have a strange theory on our hands were this not true. To see how this works, operate on \( \tilde{\psi} \) with \( \hat{\pi} \):
Aharonov-Bohm effect and geometric phase

\[ \hat{\pi} \psi = (-i\hbar \nabla - (q/c)A - (q/c)\nabla \zeta) \psi e^{i(q/\hbar c)\zeta} \]

\[ = \left\{ (-i\hbar \nabla \psi - i\hbar \psi (i(q/\hbar c)\nabla \zeta) - (q/c)A \psi - (q/c)(\nabla \zeta) \psi \right\} e^{i(q/\hbar c)\zeta} \]

\[ = \left\{ (-i\hbar \nabla - (q/c)A) \psi \right\} e^{i(q/\hbar c)\zeta} \quad (8.5) \]

The term \( \nabla \zeta \) has disappeared, and \( \zeta \) appears only in the phase factor. Applying \( \hat{\pi} \) to eqn (8.6) shows that the kinetic energy is unaffected. When this is used with the time dependent Schrödinger equation the exponential factor \( e^{i(q/\hbar c)\zeta} \) multiplies all terms in the Schrödinger equation so it cancels, and the gauge transformation does not appear. In other words, though \( A \) is changed by \( \nabla \zeta \), and the electron wave function has acquired a potentially complicated phase factor due to the gauge transformation, the observables are unaffected.

As an amusing aside, suppose you decide to simplify matters by letting \( \nabla \zeta = -A \). In this case, \( A + \nabla \zeta \) vanishes and \( p - (q/c)A \) becomes \( p \). Referring to eqn (8.5), the third and fourth terms cancel. However, the second term is now \( (-i\hbar)(iq/\hbar c)(-A) = -(q/c)A \), so nothing has changed. Nice try.

As a consequence of the gauge transformation, \( \hat{\psi} \) is related to \( \psi \) by

\[ \hat{\psi} = U \psi \]

where

\[ U = e^{i(q/\hbar c)\zeta} \quad (8.7) \]

You are undoubtedly familiar, from classical mechanics and standard quantum mechanics, with generators of transformations: \( H \) is the generator of time evolution; momentum is the generator of spatial displacement; the discrete generators of point groups; and so on. Here we have a transformation: \( U = e^{i(q/\hbar c)\zeta} \), whose generator is charge. Thus, it can be said that charge is the generator of gauge transformations. This conclusion is far-reaching, applying to charges other than the electric one considered here. Conserved currents and their charges can be described using Noether's theorem (Appendix 4).

The transformation \( U \) is multiplication by a (unit magnitude) phase factor whose argument is a function of spatial location. For each spatial location, it is a rotation in the complex plane. It is unitary and one dimensional. The gauge group is labeled U(1). This is the gauge group of quantum electrodynamics (QED). Other aspects of QED are challenging, but not its gauge group.

Covariant derivative (advanced)

The above results can be distilled into a neat algorithm. To see how this works, consider a charged particle in the presence of an electromagnetic field. We begin by assuming

\footnote{For \( U \) to have meaning in quantum mechanics it must be single valued. This requires that the space on which \( U \) operates is simply connected so the integral of \( \nabla \zeta \) depends on the end points.}
Aharonov-Bohm effect and geometric phase

that the system is expressed in terms of the charged particle and the field, but without any coupling between them. In other words, the Hamiltonian is \( p^2/2m + H_{EM} \), where \( H_{EM} \) is the electromagnetic energy that was discussed in Chapter 3. The interaction is now turned on. The electrodynamics interaction is achieved by promoting the canonical momentum from \( p^\mu = i\hbar \partial^\mu \), to a new momentum: \( i\hbar \partial^\mu - (q/c)A^\mu \). In other words,

\[
\partial^\mu \rightarrow \partial^\mu + i(q/\hbar c)A^\mu = D^\mu
\]

(8.9)

The symbol \( D^\mu \) denotes what is called a covariant derivative. You might wonder why something like \( D^\mu \) is called a covariant derivative, as it is obviously contravariant. A reasonable person might find this baffling, and rightly so.

The confusion arises because the term covariant has two meanings. To make matters worse, authors tend to mingle them without mentioning the distinction — leaving that to context. This passes muster once you know what is going on, but it can be confusing on first encounter. When an equation or expression does not change its form under a Lorentz transformation it is said to be Lorentz covariant. For example: 4-vectors like \( p^\mu \) are Lorentz covariant; Maxwell's equations and the Dirac equation are Lorentz covariant; the Schrödinger equation is not Lorentz covariant; and so on. The covariant derivative \( D^\mu \) falls into this category. However, we also deal with covariant and contravariant components of tensors. This is different. In this context, these terms denote the transformation properties of the tensor components. With this distinction in mind, we now see that \( D^\mu \) and \( D_\mu \) are each Lorentz covariant, whereas \( D^\mu \) transforms as a contravariant 4-vector, while \( D_\mu \) transforms as a covariant 4-vector.

The covariant derivative converts the system from one of global gauge invariance to one of local gauge invariance, and when it undergoes a Lorentz transformation it mixes electric and magnetic interactions according to the requirements of special relativity. Without it the system is not invariant with respect to a local gauge transformation. Local gauge invariance is assured through the substitution given by eqn (8.9). This is referred to as the minimal substitution because it is the least complicated way to ensure gauge invariance. In other words, without the gauge field \( A^\mu \) there is no straightforward mechanism for ensuring gauge invariance. This idea was illustrated in Fig. 7, where it was pointed out that if \( A^\mu \) changes, so must the wave function's phase factor.

When quantization of the electromagnetic field was examined in the last chapter, we saw how gauge transformations work in electromagnetic theory. To further explore how gauge invariance works in the quantum mechanics of particles interacting with fields, let us revisit the material surrounding eqns (8.4) – (8.8). Namely, a gauge transformation is applied to the vector potential \( A \) by adding the gradient of a scalar: \( A \rightarrow A + \nabla \zeta \), while at the same time the wave function undergoes: \( \psi \rightarrow \psi e^{i(q/\hbar c)\zeta} \). These are called gauge transformations of the second and first kind, respectively — terms that were introduced
by Wolfgang Pauli. The easiest way to handle manipulations is by using the unitary transformation given by eqn (8.8): \( U = e^{i(q/c)\zeta} \). Let us start by applying this transformation to the kinetic momentum: \( \pi = p - (q/c)A \), where \( p = -ih\nabla \):

\[
\tilde{\pi} = U\pi U^\dagger = U \left( p - (q/c)A \right) U^\dagger = UpU^\dagger - (q/c)A
\]

(8.10)

\[
= p - (q/c)A - (q/c)\nabla \zeta
\]

(8.11)

Obviously, \( \pi \) is recovered from \( \tilde{\pi} \) by using \( \pi = U^\dagger \tilde{\pi} U \). Now the Schrödinger equation is transformed from one involving \( H \) and \( \psi \) to one involving \( \tilde{H} \) and \( \tilde{\psi} \) by operating from the left with \( U \):

\[
H\psi = i\hbar \frac{\partial \psi}{\partial t} \quad \Rightarrow \quad UHU^\dagger U\psi = i\hbar \frac{\partial U\psi}{\partial t} \quad \Rightarrow \quad \tilde{H}\tilde{\psi} = i\hbar \frac{\partial \tilde{\psi}}{\partial t}
\]

(8.12)

where \( \tilde{H} = UHU^\dagger \) is

\[
\tilde{H} = \frac{1}{2m} \left\{ U\pi U^\dagger \cdot U\pi U^\dagger \right\} = \frac{1}{2m} \left\{ U \left( p - (q/c)A \right) U^\dagger \cdot U \left( p - (q/c)A \right) U^\dagger \right\}
\]

(8.13)

\[
= \frac{1}{2m} \left\{ \left( p - (q/c)A - (q/c)\nabla \zeta \right) \cdot \left( p - (q/c)A - (q/c)\nabla \zeta \right) \right\} = \frac{\tilde{\pi} \cdot \tilde{\pi}}{2m}
\]

(8.14)

Because \( \tilde{H}\tilde{\psi} = i\hbar \frac{\partial \tilde{\psi}}{\partial t} \) is the same as \( H\psi = i\hbar \frac{\partial \psi}{\partial t} \), each form gives identical physical results. These manipulations have been carried out using the vector potential \( A \), with the electric potential \( \phi \) set to zero. Now let's see how things work using \( \phi \neq 0 \). With \( A = 0 \) and \( \phi \neq 0 \), apply the gauge transformations: \( \phi \rightarrow \phi + \phi^0 \) and \( U = \exp \left( -i(q/c)\int \! dt \phi^0 \right) \) to the Schrödinger equation:

\[
\tilde{H}\tilde{\psi} = \left( \frac{p^2}{2m} + q\phi + q\phi^0 \right)\tilde{\psi} = i\hbar \frac{\partial \tilde{\psi}}{\partial t} = i\hbar \frac{\partial}{\partial t} \left( \psi \exp \left( -i(q/c)\int \! dt \phi^0 \right) \right)
\]

(8.15)

\[
= \left( i\hbar \frac{\partial \psi}{\partial t} + q\phi^0 \psi \right) \exp \left( -i(q/c)\int \! dt \phi^0 \right)
\]

The exponential multiplies each of the terms, so it cancels, and the \( q\phi^0 \) terms also cancel. This leaves

\[
\left( \frac{p^2}{2m} + q\phi \right)\psi = i\hbar \frac{\partial \psi}{\partial t}
\]

(8.16)

Again, we see that \( \tilde{H}\tilde{\psi} = i\hbar \frac{\partial \tilde{\psi}}{\partial t} \) is the same as \( H\psi = i\hbar \frac{\partial \psi}{\partial t} \). Referring to eqn (8.3), it is now clear how the 4-vector form works. Expressing it in terms of its \( \phi \) and \( A \) parts:
Aharonov-Bohm effect and geometric phase

\[
\exp\left(-i\frac{q}{\hbar c} \int dx^\mu A_\mu\right) = \exp\left(-i\frac{(q/\hbar)\int dt \phi + i(q/\hbar c)\int d\mathbf{r} \cdot \mathbf{A}}{2}\right)
\]  
(8.17)

**Local gauge invariance (advanced)**

Gauge invariance is now examined from a different perspective. Namely, it is the *requirement of local gauge invariance* that leads inevitably to the minimal substitution \( p \to p - (q/c)A \) and all that follows from it. In other words, the requirement of local gauge invariance ensures (demands) that the charged particle and the electromagnetic field couple to one another. In fact, it even dictates the form that this coupling must assume. From this perspective, the requirement of local gauge invariance is the driving force, and quantum electrodynamics is a consequence. Do not underestimate the depth and utility of this gauge principle. It embodies an essential interplay between classical and quantum physics: Classically when \( A \) and \( \phi \) appear they can be converted to expressions containing \( E \) and \( B \). Quantum mechanically this cannot be achieved because phase enters in a way that has no classical counterpart. Without local gauge invariance quantum mechanics is inconsistent.

The reason for assigning this privileged status to the gauge principle can be illustrated using the free particle Schrödinger equation. Let’s see how this works by considering a particle in what we at first believe is field-free space, and demand that quantum mechanics satisfies local gauge invariance. This will reveal why and how the electromagnetic field must be included in the equations of motion.

A solution of the Schrödinger equation yields a wave function whose phase is immaterial when it comes to measurement. Despite the fact that we routinely use complex functions like \( e^{ikx} \), measurements tell us about the magnitude \( |\psi| \) rather than \( \psi \). Phase information is lost. Everyone knows that multiplication of a total wave function by \( e^{i\alpha} \), where \( \alpha \) is a real constant, has no affect. This is an example of global phase. Invariance with respect to a global phase (gauge) transformation gives rise to conservation of charge. Let’s now promote the global phase \( \alpha \), which is independent of space and time, to \( \alpha(r,t) \), which depends on space and time. Though \( \psi \to \psi e^{i\alpha} \) is now \( \psi \to \psi e^{i\alpha(r,t)} \), the magnitude of \( \psi \) is unaffected, because \( |\psi e^{i\alpha}| = |\psi e^{i\alpha(r,t)}| \). Applying this gauge transformation to the wave function, the Schrödinger equation is

\[
i\hbar \frac{\partial}{\partial t}(\psi e^{i\alpha(r,t)}) = \frac{1}{2m} (-i\hbar \nabla)^2 \psi e^{i\alpha(r,t)}
\]

(8.18)

Hereafter, \( \alpha \) will be understood to mean \( \alpha(r,t) \). A bit of algebra yields

\[
i\hbar \frac{\partial \psi}{\partial t} = \left\{ \hbar \frac{\partial \alpha}{\partial t} + \frac{1}{2m} (-i\hbar \nabla + \hbar \alpha)^2 \right\} \psi
\]

(8.19)
The only way that this expression can recover the original Schrödinger equation is to alter the original Schrödinger equation. Specifically, the original Schrödinger equation must contain one or more items that enable cancellation of the terms $\hbar \nabla \alpha$ and $\hbar \partial \alpha / \partial t$ that have entered through differentiation of the gauge transformed wave function. In other words, the term $-i\hbar \nabla$ in eqn (8.18) must be supplemented in a way that enables $\hbar \nabla \alpha$ and $\hbar \partial \alpha / \partial t$ in eqn (8.19) to be eliminated.

This is achieved, in part, by including in the Schrödinger equation a vector field that can have $-\hbar \nabla \alpha$ added to it, and in so doing cancel the $\hbar \nabla \alpha$ in eqn (8.19). The presence of this vector field at the outset can be exploited to ensure that no physical effect arises in the equations because of the gauge transformation applied to $\psi$. Let us call this vector field $-(q/c)A$. In other words, we are now starting with $-i\hbar \nabla -(q/c)A$ in eqn (8.18) instead of just $-i\hbar \nabla$. Consequently, we are also starting with $(-i\hbar \nabla - (q/c)A + \hbar \nabla \alpha)$ in eqn (8.19) instead of just $(-i\hbar \nabla + \hbar \nabla \alpha)$. If the quantity $(hc/q)\nabla \alpha$ is now added to $A$, all is well:

$$-i\hbar \nabla - (q/c)A + \hbar \nabla \alpha \Rightarrow -i\hbar \nabla - (q/c)(A + (hc/q)\nabla \alpha) + \hbar \nabla \alpha = -i\hbar \nabla - (q/c)A \quad (8.20)$$

This shows that a local gauge transformation of the wave function can be used as long as it is accompanied by a change of the gauge field: $A \rightarrow A + (hc/q)\nabla \alpha$. Consistency with previous notation is achieved by using $\alpha = (q/hc)\zeta$. In other words, $\psi \rightarrow \psi e^{i(q/hc)\zeta}$ must be accompanied by $A \rightarrow A + \nabla \zeta$.

Application of the gauge principle has resulted in a requirement that a vector field be present in the quantum mechanical equations of motion. Furthermore, the requirement that no physical consequence results from the transformation: $A \rightarrow A + \nabla \zeta$ implies that, for the classical counterpart, the physical content of $A$ is manifest in its curl. But this is nothing more that electromagnetism: $B = \nabla \times A$.

The term $\hbar \nabla \alpha$ in eqn (8.19) has now been eliminated, but $\hbar \partial \alpha / \partial t$ in eqn (8.19) must also be eliminated. On the basis of the above introduction of the vector field $A$, the manner in which this works can be anticipated. Namely, if the original Schrödinger equation contains the potential energy $q\phi$, this needs to undergo the transformation $\phi \rightarrow \phi - \partial \zeta / \partial t$ in order to cancel the term $\hbar \partial \alpha / \partial t$ [which is equal to $(q/c)\partial \zeta / \partial t$] in eqn (8.19). This leaves the electric field unchanged:

$$E = -\nabla \phi - \partial A / \partial t \Rightarrow -\nabla (\phi - \partial \zeta / \partial t) - \partial (A + \nabla \zeta) / \partial t$$

$$= -\nabla \phi + \nabla \partial \zeta / \partial t - \partial A / \partial t - \partial \nabla \zeta / \partial t \quad (8.21)$$

A complete picture is now in place. The requirement of local gauge invariance implies the presence of the vector and scalar quantities $A$ and $\phi$ (equivalently, $A^\mu$). These are in fact the electromagnetic potentials. The gauge invariant Schrödinger equation is

$$i\hbar \frac{\partial \psi}{\partial t} = \left( \frac{1}{2m} (-i\hbar \nabla - (q/c)A)^2 + q\phi \right) \psi \quad (8.22)$$
Aharonov-Bohm effect and geometric phase

The role of electric charge is subtle. It is interesting that \( q \) multiplies the fields \( A \) and \( \phi \), given that \( q \) is a property of the particle. For example, the interaction of \( A \) with the charged particle is already partly in place with \( \langle q/c \rangle A \). The dynamical part of the coupling arises from \( p \) interacting with \( \langle q/c \rangle A \). Thus, terms like \( p \cdot \langle q/c \rangle A \) should not be thought of simply as a charged particle (described by \( p \)) interacting with a field \( \langle q/c \rangle A \).

The 4-vector \( A^\mu \) is called a gauge field because it ensures local gauge invariance. We were able to work through the AB effect without invoking the gauge principle because the math is straightforward. However, life is not as easy with the weak and strong forces, where the gauge principle is invaluable. We will not go into these topics. However, we shall see that use of the gauge principle with the Born-Oppenheimer approximation offers great advantage.

Local gauge invariance: the gauge principle

The transformation: \( \psi \rightarrow \tilde{\psi} = \psi e^{i\langle q/c \rangle \zeta} \)

implies \( \langle q/c \rangle A \) and \( q\phi \), i.e. the gauge field \( A^\mu \).

\[
A \rightarrow \tilde{A} = A + \nabla \zeta \quad \phi \rightarrow \tilde{\phi} = \phi - \partial \zeta / \partial t
\]

9. Variation on the main theme

We have seen that an electron that passes through an \( A \neq 0 \) region acquires a phase: \( \gamma = -(e/\hbar c) \int \mathbf{d}r \cdot \mathbf{A} \). Though the highly idealized situation indicated in Fig. 5 enlists 1D (albeit along a curved trajectory) monoenergetic waves of uniform magnitude, the same phase would be appended to 1D wave packets, because each of its \( k \)-components acquires the same phase.

Electron in a box

Taking this a step further, suppose an electron is contained inside a hollow box and this box-plus-electron combination is transported around the solenoid. A stationary box has a set of single particle energies \( E_n \) and eigenfunctions \( \psi_n \). For a box whose dimensions are sufficiently small, the energies \( E_n \) will be large relative to any influence that \( A \) might have on them. Likewise, as long as the box is transported slowly, the eigenvalues and eigenfunctions of the stationary box are expected to hold, to a good degree of accuracy, in the absence of degeneracy. This is the regime of interest. It is assumed that the box has sides whose lengths differ such that there is no degeneracy, and that the electron is in one of the eigenstates of the box.

As mentioned above, as long as the box is transported slowly, it is reasonable to expect the solutions for the stationary box to remain good approximations. For example, the mo-
tion of the box will distort ever so slightly the electron wave functions $\psi_n$, causing them to take on a mixed character in the basis of stationary box functions. However, as long as the energy separation between levels remains large compared to any energy imparted to the particle due to motion of the box, this is a minor effect. The electron wave functions acquire the phase $\gamma$ as the box is transported, because this depends only on the line integral of $A$ along the path taken by the box.

Now consider this from the perspective of an adiabatic separation. The environment inside the box has associated with it a characteristic time scale that is dictated by the box dimensions and the particle mass (in this case that of an electron). This time scale is rapid relative to the one on which external parameters are varied. The large difference in time scales justifies the adiabatic approximation. Electron dynamics inside the box average out insofar as their contribution to the phase $\gamma$ is concerned. The only relevant coordinate for computing the phase is the location of the box.

(Add periodic potential stuff.)

10. Born-Oppenheimer approximation and local gauge invariance

The gauge principle is now applied to Born-Oppenheimer states of polyatomic molecules. There is an important distinction between this case and the examples in the previous section. There, control of the external parameters was absolute. For example, the electron-in-a-box follows adiabatically the slow variation of the box location without influencing it in turn. There is no wave function for the external parameters, as they are completely under our control.

With the Born Oppenheimer approximation (BOA), the Schrödinger equation for the nuclear degrees of freedom can be solved for each of the adiabatic potential energy surfaces (adiabats). The BOA is great for solving the electron problem and then obtaining vibrational wave functions. Its breakdown is taken into account by computing couplings due to terms that are omitted from the Hamiltonian in making the BOA. We shall see that the BOA also leads to an interesting phase. This phase is the object of our attention in this section and the next.

Phase and flux

Momentum operators for nuclear degrees of freedom play a central role in nonadiabatic dynamics because nuclear motions couple adiabats. To express these operators in a Schrödinger equation for the nuclei, it is necessary to integrate over the electron coordinates. For example, the action of $p = -i\hbar \nabla$ on $\chi\psi$ (where $\chi$ and $\psi$ are nuclear and electron wave functions, respectively, and it is understood that $\nabla$ operates only with respect to nuclear coordinates) is multiplied from the left by $\psi$, and integration is carried out over the electron coordinates. Thus, the effective nuclear kinetic momentum operator, $\pi$, is

$$\pi \chi = \langle \psi | p \chi \psi \rangle$$

(10.1)
Keep in mind that $\psi$ depends parametrically on the nuclear coordinates.

Despite the fact that the wave function $\chi\psi$ is single valued, $\chi$ and $\psi$ are not in general separately single valued, though we may wish to make them so for convenience. For example, a gauge transformation of the electron wave function: $\psi \rightarrow \tilde{\psi} = \psi e^{i\eta}$ (where $\eta$ is a function of external parameters $R$, which here are the nuclear coordinates) can be used to obtain a single valued $\tilde{\psi}$. Note that a gauge transformation applied to the electron wave function must be manifest in $\tilde{\chi}$ to ensure that $\tilde{\chi}\tilde{\psi}$ is single valued. In other words, $\tilde{\psi} = \psi e^{i\eta}$ requires $\tilde{\chi} = \chi e^{-i\eta}$. In addition, it is a good idea to keep track of terms that account for transitions, i.e., $\langle \psi | p \chi \psi \rangle$, as these play a central role in the theory of nonadiabatic transitions. All of this is well known.

Let us now return to local gauge invariance. Any time a wave function is subjected to a local gauge transformation, a vector field must be present that enables the affect on the dynamics incurred through this transformation on the wave function to be counteracted. In other words, the overall gauge transformation (i.e., of the wave function and the vector field) must not result in additional physics. This was discussed in Section 8.

In the present case, the requirement of local gauge invariance implies the existence of a vector field that we shall call $f$. As in Section 8, this vector field is present in such a way that it can have something added to it to ensure that no physical effect enters the mathematics via the gauge transformation. The something that needs to be added is the gradient of a scalar because this is integrable. That is, because $dR \cdot \nabla \eta = i \lambda$, the integral of $dR \cdot \nabla \eta$ depends only on the end points, and therefore no flux is enclosed in a circuit integral of $dR \cdot \nabla \eta$. The circuit integral of the gauge field in general encloses flux, whereas the circuit integral of $dR \cdot \nabla \eta$ does not. This is how gauge invariance is achieved. This procedure was used in Section 8, where the gauge principle was applied to electrodynamics.

In electrodynamics, the charge of the particle is coupled to the gauge field $A^\mu$, and the momentum of the particle enters through terms like $p \cdot (q/c)A$. In the Born-Oppenheimer case, the gauge field must likewise be proportional to a momentum in order that its line integral yields an action. This gauge field can be obtained by considering how nuclear momentum is coupled to electron degrees of freedom, of course averaged over electron density. Thus, eqn (10.1) identifies the gauge field: $f = \langle \psi | p \psi \rangle$, through the operation:

$$\pi \chi = \langle \psi | p \chi \psi \rangle = \left( p + \langle \psi | p \psi \rangle \right) \chi = f$$  \hspace{1cm} (10.2)

Invariance with respect to gauge transformation is verified by the substitutions: $\tilde{\chi} = \chi e^{-i\eta}$ and $\tilde{f} = f + h\nabla \eta$. This differs slightly from the phase used to gauge transform

---

\textsuperscript{4} To see how this works, recall electromagnetism: $\oint d\ell \cdot A = \int dS \cdot \nabla \times A = \int dS \cdot B = \Phi$. The addition of $\nabla \eta$ to $A$ results in the term: $\oint d\ell \cdot \nabla \eta$. However, this vanishes because $\nabla \times \nabla \eta = 0$. Therefore the circuit integral of $\nabla \eta$ does not enclose flux.

\textsuperscript{\textdagger} $\pi \tilde{\chi} = (p + f + h\nabla \eta) \chi e^{-i\eta} = \left( \left( p + (-i\hbar)(-i\nabla \eta) + f + h\nabla \eta \right) \chi \right) e^{-i\eta}$. The two $\nabla \eta$ terms inside the large parentheses cancel, leaving $\left( (p + f) \chi \right) e^{-i\eta}$. 231
Aharonov-Bohm effect and geometric phase

the wave function in the AB case. There, $e^{i(q/\hbar)\chi}$ was used, with the sign in the exponent taken to be positive, whereas here (with $e^{-i\eta}$) the sign in the exponent is taken to be negative. The reason for using $e^{-i\eta}$ here is because $\chi$ describes nuclei. In other words, $\tilde{\chi} = \chi e^{-i\eta}$ implies for the corresponding gauge transformed electron wave function $\tilde{\psi} = \psi e^{i\eta}$ to ensure single valuedness of $\tilde{\chi}\tilde{\psi}$.

The desired phase is obtained by closed circuit integration of the gauge field. With the AB effect, the phase $\beta$ was obtained through integration of $(q/c)A$ over a path, then dividing the result by $\hbar$. Here, the gauge field is $f$, so the integration is

$$\gamma(C) = \frac{1}{c} \oint_C dR \cdot f / \hbar = i \frac{1}{c} \oint_C dR \cdot \langle \psi | \nabla \psi \rangle$$

(10.3)

Applying Stokes' theorem yields

$$\gamma(C) = i \int_{S_C} dS \cdot \nabla \times \langle \psi | \nabla \psi \rangle$$

(10.4)

With $\gamma(C)$ expressed as a surface integral, it is clear that it is invariant with respect to the transformation: $f \rightarrow f + \hbar \nabla \eta$. In other words, $\nabla \times (f + \hbar \nabla \eta) = \nabla \times f$ because the curl of the gradient of a scalar is identically zero.

Geometry issues loom. Stokes' theorem presupposes a 3D $R$-space. In general, however, the number of participating nuclear degrees of freedom is not limited to three. To deal with spaces having more than three dimensions, differential geometry is used, as this generalizes Stokes' theorem to an arbitrary number of dimensions. This works nicely, but if you are not familiar with differential geometry, it will take a while to acquire a useful level of facility. From eqn (10.4), we see that $\gamma(C)$ is a flux through the surface enclosed by $C$ (Fig. 11) in the space of the nuclear coordinates that account for the adiabatic evolution of the wave function. The phase $\gamma(C)$ that we have obtained via the AB effect and local gauge invariance is called the geometric phase.

![Figure 11](image)

The expression for the geometric phase given by eqn (10.4) is in terms of a single adiabat $\psi$. However, the presence of one or more additional adiabats is implicit in the gradient of $\psi$, namely, the fact that $\nabla \psi \neq 0$. In other words, suppose $\psi$ is composed of two diabats, and this composition is a function of the nuclear coordinates. Then $\nabla \psi$ will in general be nonzero. However, the fact that $\psi$ is composed of two diabats implies that there is another adiabat that is composed of the same two diabats. This marks a distinction with electrodynamics. There the gauge group is U(1). There are no transitions between electron states.
Recall the expression that was derived earlier [eqn (7.3)] for the AB phase:

$$\beta = \left(\frac{q}{\hbar c}\right) \oint dr \cdot A$$  \hfill (10.5)

The fact that $\gamma(C)$ has the same form as $\beta$ should come as no surprise, as these phases arise in the same way. In electrodynamics, charge $q$ combines with the vector field $A^\mu$ to form a momentum.\(^\text{2}\) In the present case, the nuclei undergo motion in the field presented by the electrons. If the electron wave function does not vary with the nuclear coordinate $s$, there is no coupling and the geometric phase vanishes. It is the variation of $\psi$ with $R$ that constitutes the gauge field.

### Surface integral

To evaluate the surface integral of $\nabla \times \langle \psi | \nabla \psi \rangle$ in eqn (10.4), the operator $\nabla \times$ needs to be expressed on a basis. Using the identity: $\nabla \times (h \nabla g) = (\nabla h) \times (\nabla g)$, where $h$ and $g$ are scalars, and choosing the $n$th adiabat yields $\nabla \times \langle n | \nabla n \rangle = \langle \nabla n | \nabla n \rangle$. Adding closure enables eqn (10.4) to be written:

$$\gamma_n(C) = \left(\frac{\hbar}{\sqrt{2\pi}}\right) \int dS \cdot \sum_{m \neq n} \langle n | \nabla m \rangle \times \langle m | \nabla n \rangle$$ \hfill (10.6)

The $m = n$ term is excluded because $\langle n | \nabla n \rangle \times \langle n | n \rangle = 0$. Note also that $\langle n | \nabla n \rangle$ is imaginary, so $\langle n | \nabla n \rangle \times \langle n | \nabla n \rangle$ is real and $\gamma_n(C)$ must be real. A common maneuver is to express $\langle m | \nabla n \rangle$ in terms of gradients of $H$ by taking the gradient of $H | n \rangle = E_n | n \rangle$ and operating from the left with $\langle m |$. This yields

$$\langle m | \nabla H | n \rangle + \langle m | H | \nabla n \rangle = \langle m | \nabla E_n | n \rangle + \langle m | E_n | \nabla n \rangle$$ \hfill (10.7)

In general, $E_n$ can be expressed as an explicit function of nuclear coordinates. However, we are dealing with a closed circuit, and no harm is done in choosing a path for which $E_n$ remains constant. In this case, the first term on the right hand side vanishes, leaving

$$\langle m | \nabla n \rangle = \frac{\langle m | \nabla H | n \rangle}{E_n - E_m} \quad \text{where } m \neq n$$ \hfill (10.8)

Using this with eqn (10.6) yields

$$\gamma_n(C) = -\oint_{c_C} dS \cdot V_n \quad V_n = \text{Im} \sum_{m \neq n} \frac{\langle n | \nabla H | m \rangle \times \langle m | \nabla H | n \rangle}{(E_n - E_m)^2}$$ \hfill (10.9)

\(^2\) Keep in mind that this momentum is quite different from the momentum of an electromagnetic field: $E \times B / c^2$.
An advantage of eqn (10.9) over the path integral is that it is relatively easy to compute gradients of the Hamiltonian and evaluate the matrix elements.

**Two states and degeneracy**

The geometric nature of $\gamma(C)$ has emerged, e.g., as illustrated in Fig. 11. Manipulations that further highlight the geometric nature of $\gamma(C)$ are now carried out. To begin, consider two states and their $\mathbf{R} = 0$ degeneracy, e.g., two intersecting adiabats. Because only two states are involved, the sum in eqn (10.9) reduces to one term:

$$V_n = \text{Im} \left( \frac{\langle n | \nabla H | m \rangle \times \langle m | \nabla H | n \rangle}{(E_n - E_m)^2} \right)$$

(10.10)

A form for the Hamiltonian is used that can be evaluated readily using eqn (10.10). The most general form for a $2 \times 2$ Hamiltonian matrix is

$$H = \frac{1}{2} \sigma \cdot \mathbf{R} = \frac{1}{2} (\sigma_1 X + \sigma_2 Y + \sigma_3 Z) = \frac{1}{2} \begin{bmatrix} Z & X - iY \\ X + iY & -Z \end{bmatrix}$$

(10.11)

The $\sigma_i$ are Pauli matrices, the parameters $X$, $Y$, and $Z$ are real, and the factor of $\frac{1}{2}$ is there because it turns out to be convenient. Energies are relative to the $\mathbf{R} = 0$ degeneracy, so there is no need to include a $2 \times 2$ unit matrix. The eigenvalues of the matrix in eqn (10.11) (including the $\frac{1}{2}$) are $\pm R/2$, where $R^2 = X^2 + Y^2 + Z^2$. The parameter space vector $\mathbf{R}$ changes slowly, causing the eigenvalues and eigenvectors to evolve adiabatically (Fig. 12).

![Figure 12](image)

You might wonder why the form used in eqn (10.11) has been chosen. What insight prompted its enlistment? The answer is that when two states are expressed in a two-state basis, with a degeneracy at $\mathbf{R} = 0$, the system has a spin-$\frac{1}{2}$ representation. Consequently, it is known *a priori* that the wave function undergoes a sign reversal in $2\pi$. This corresponds to $Y = 0$ in eqn (10.11). Our two state model with $Y \neq 0$ is more general so let’s see what happens.
Aharonov-Bohm effect and geometric phase

Referring to Fig. 12(c), the natural quantization direction for $\sigma$ is $R$: as $R$ varies $\sigma$ follows. The matrix element in eqn (10.10), is evaluated using the gradient of $H = \sigma \cdot R / 2$, i.e., $\nabla H = \sigma / 2$. The notation $|n\rangle = |+\rangle$ and $|m\rangle = |\rangle$ is introduced in deference to the spin representation. The states follow adiabatically the evolution of $R$, as indicated in Fig. 12(c). Because $\sigma$ is referenced to $R$, the matrix element in eqn (10.10) is easily evaluated:

$$\langle + | \nabla H | - \rangle = \frac{1}{2} (| + \rangle \langle \sigma_1 | - \rangle \hat{e}_1 + | + \rangle \langle \sigma_2 | - \rangle \hat{e}_2 + | + \rangle \langle \sigma_3 | - \rangle \hat{e}_3) = \frac{1}{2} (\hat{e}_1 - i \hat{e}_2)$$

(10.12)

Primes indicate a reference system that retains its orientation with respect to $R$ throughout the evolution. The above matrix element was evaluated using

$$| - \rangle = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad | + \rangle = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$$

(10.13)

The matrix element $\langle + | \nabla H | - \rangle$ is put into eqn (10.10), and $E_+ - E_- = R$ is used to write:

$$V_+ = \text{Im} \left( \frac{(\hat{e}_1 - i \hat{e}_2) \times (\hat{e}_1 + i \hat{e}_2)}{4R^2} \right) = \frac{\hat{e}_3}{2R^2}$$

(10.14)

Thus, $V_+$ points in the $R$ direction, which is, by definition, $\hat{e}_3$. As $R$ evolves, its tip traces a path, returning to its original position. Using eqn (10.14) with eqn (10.9) gives

$$\gamma_+(C) = - \int_{S_c} dS \cdot \frac{\hat{e}_3}{2R^2} = - \int_{S_c} R^2 d\Omega \cdot \frac{\hat{e}_3}{2R^2} = - \frac{1}{2} \Omega$$

(10.15)

Clearly $\gamma_-(C) = + \frac{1}{2} \Omega$. Thus, the phase for the closed circuit $C$ is

$$\gamma_+ (C) = \pm \frac{1}{2} \Omega$$

(10.16)

The solid angle $\Omega$ subtended from the $R = 0$ degeneracy point (Fig. 13) can take on a continuous range of values between 0 and $2\pi$, dictated by the path of $R$. When the matrix elements of $H$ are real (i.e., $Y = 0$), $R$ completes its circuit in the $\hat{e}_3 \hat{e}_1$ plane. Because this plane contains $R = 0$, it follows that $\gamma_+ (C) = \pm \pi$. This is a conical intersection. For the general case of a complex Hamiltonian matrix, an infinite number of planes contain $R = 0$. This is relevant to molecules having an odd number of electrons.

Figure 13. The closed circuit $C$ defines a solid angle $\Omega$ in a 3D $R$-space.
Spin in a magnetic field

In light of the connection that has just been made to spin, let us now consider the case of an arbitrary spin in a magnetic field. If $B$ slowly changes its laboratory direction such that it traces out a path that returns it to its original orientation (Fig. 14), the Hamiltonian varies accordingly. The spin eigenstates follow adiabatically, remaining referenced to $B$ throughout its slow evolution. Upon completion of the path, $B$ has returned to its original orientation, but has the spin state also recovered its original form, or has it acquired a phase due to its adiabatic passage around the path? From the previous section, we know how spin behaves, so we now examine integer or odd-half-integer spin interacting with $B$. The gradient of $H = KS \cdot B$ is $KS$, and the energies are $KBm$. Thus, eqn (10.9) for a given $m$ is

$$V_m = \Im \sum_{m' \neq m} \frac{\langle m | S | m' \rangle \times \langle m' | S | m \rangle}{B^2 (m - m')^2}$$  \hspace{1cm} (10.17)

The use of a double prime avoids confusion with the rotated axes, which are labeled with a single prime. The matrix elements are easily evaluated:

$$\langle m | S | m' \rangle = \langle m | S_{1'} \hat{e}_1 + S_{2'} \hat{e}_2 + S_{3'} \hat{e}_3 | m' \rangle = \langle m | S_{1'} | m \pm 1 \rangle \hat{e}_1 + \langle m | S_{2'} | m \pm 1 \rangle \hat{e}_2$$

$$= \frac{1}{2} \left( \langle m | S^+ + S^- | m \pm 1 \rangle \hat{e}_1 - i \langle m | S^+ - S^- | m \pm 1 \rangle \hat{e}_2 \right)$$  \hspace{1cm} (10.18)

Nonzero matrix elements have $m'' = m \pm 1$. Some algebra yields $V_m = (m/B^2) \hat{e}_3$. As with eqn (10.15), integration over $dS$ is trivial because $V_m$ is parallel to $S$. For a surface enclosed by $C$, integration yields the geometric phase:

$$\gamma_m(C) = -m \Omega$$  \hspace{1cm} (10.19)

This differs from eqn (10.16) in that $\mp \frac{1}{2}$ has been replaced by $-m$. When $m$ is an integer and $\Omega$ is $2\pi$, $\gamma_m(C)$ is an integer multiple of $2\pi$. In this case the wave function does not change sign when $\Omega = 2\pi$. On the other hand, when $m$ is odd-half-integer, the wave function changes sign when $\Omega = 2\pi$. 

![Figure 14. B has constant magnitude as its direction traces a closed path. The tip of the thick arrow moves on the surface of a sphere.](image-url)
11. Geometric phase

We have seen that one of the ways in which adiabatic change is manifest is through the geometric phase it engenders. The origins of what nowadays falls under the heading of geometric phase go back a long way: a theorem due to Wigner and Von Neumann, the Jahn-Teller effect (Peierls distortion), the Aharonov-Bohm effect, flux quantization in superconductivity, Pancharatnam and Berry phases, etc. Geometric phase serves as an essential ingredient in theories of intersecting potential surfaces and associated nonadiabatic dynamics. Appendix 6 describes the classical geometric phase that arises from the parallel transport of a vector on the surface of a sphere, e.g., as with the Foucault pendulum. You might find this interesting. Quantum mechanically, adiabatic evolution means that the state does not change its quantum numbers. For example, nodes in a wave function cannot be created or annihilated, as this would require precipitous change of the curvature of the wave function.

In this section, the approach introduced by Michael Berry will be presented. Many of the geometric phase results derived by Berry have already been obtained via local gauge invariance. The close relationships among the AB effect (and its progenitors), gauge field theory, and the material in this section are stunning. It is amusing to note the uncanny resemblance between Michael Berry and Reinhard Schinke.

General situation

A state whose ket is \( |n\rangle \) evolves adiabatically under the influence of external parameters that we shall denote collectively as \( R \). These parameters vary on a time scale that is slow relative to the one on which the particle dynamics associated with \( |n\rangle \) transpire. As \( R \) undergoes its slow evolution, the time independent Schrödinger equation yields eigenvalues and eigenfunctions at each instant of time, like a series of snapshots:

\[
H(R)|n(R)\rangle = E_n(R)|n(R)\rangle
\]  

Equation (11.1) yields a separate set of eigenvalues/eigenkets for each value of \( R \). The standard procedure of electronic structure theory is to solve eqn (11.1) at different \( R \). Consequently, this results in there being no relationship between the phases of the solutions at different \( R \). We are inclined to think that these unknown phases, if they vary, do so smoothly with respect to \( R \). However, eqn (11.1) does not say anything per se about relative phase. After all, it is solved for one value of \( R \), then for another, and so on. In addition, we might want \( |n(R)\rangle \) to be single valued and differentiable (except perhaps at the point of closing a circuit), because it will be necessary to take derivatives with respect to the parameters that comprise \( R \) as a path in \( R \)-space is followed.
To have (nonrelativistic) quantum mechanics, the system must satisfy a Schrödinger equation, or something close to it, at all points in time throughout the adiabatic evolution:

$$H(R)|\psi_n\rangle = i \frac{d}{dt}|\psi_n\rangle$$

(11.2)

The use of a total derivative on the right side is necessary in order to account for the adiabatic evolution. It turns out that the geometric phase does not depend on the amount of time required to complete the adiabatic cycle. However, it is necessary to take into account the changes that transpire in the space of parameters $R$, and this is the reason for using the total derivative instead of a partial derivative in eqn (11.2).

The state vector $|\psi_n\rangle$ includes phase factors for the usual Schrödinger eigenstate phase, as well as an additional phase $\gamma_n(R)$:

$$|\psi_n\rangle = \exp\left(-i \int_0^t E_n(R) dt \right) e^{i\gamma_n(R)}|n(R)\rangle$$

(11.3)

The phase $\gamma_n(R)$ (hereafter denoted simply $\gamma_n$) is due to the adiabatic evolution of $R$. The rest is standard nonrelativistic quantum mechanics. The ansatz given by eqn (11.3) was used by Berry in his 1984 paper. The term $\gamma_n$ must be present. It is an admission of our ignorance regarding how the phase evolves with $R$.

Equation (11.3) can also be understood through the requirement of local gauge invariance. This results in $\gamma_n$ automatically having the form of an integral of a gauge field on the parameter space. We have seen that a local gauge transformation dictates the form of the gauge field. Also, because phase evolves on the parameter space, we see that only $R$ affects the phase. The phase varies according to changes in $R$, not how long it takes these changes to happen. For example, if $R$ remains constant over a long stretch of time, $\gamma_n$ does not change at all.

An equation for $\gamma_n$ is obtained by putting the $|\psi_n\rangle$ given by eqn (11.3) into the right hand side of eqn (11.2). This yields three terms on the right hand side, one of which cancels the term on the left hand side. A little algebra yields

$$\frac{d\gamma_n}{dt}|\psi_n\rangle = i \exp\left(-i \int_0^t E_n dt \right) e^{i\gamma_n} \frac{d}{dt}|n\rangle$$

(11.4)

The gradient is with respect to the parameters that comprise the $R$-space. Now operate on eqn (11.4) from the left with $\langle \psi_n |$ and use $\langle \psi_n | \psi_n \rangle = 1$ on the left hand side to write

$$\frac{d\gamma_n}{dt} = i \langle \psi_n | \exp\left(-i \int_0^t E_n dt \right) e^{i\gamma_n} \left[\nabla n\right] \frac{dR}{dt}$$

(11.5)
Aharonov-Bohm effect and geometric phase

Using eqn (11.3) yields

$$\frac{d\gamma_n}{dt} = i \langle n | \nabla n \rangle \cdot \frac{dR}{dt} \Rightarrow d\gamma_n = i \langle n | \nabla n \rangle \cdot dR$$ (11.6)

Thus, time does not play an explicit role in determining the phase. Namely, integration of eqn (11.6) is along a path in parameter space:

$$\gamma_n = i \int_{R_0}^{R} dR \cdot \langle n | \nabla n \rangle$$ (11.7)

It is understood that $\gamma_n$ is equal to the difference between the initial and final values, i.e., $\gamma_n = \gamma_n(R) - \gamma_n(R_0)$.

Equation (11.7) shows that $\gamma_n$ depends on the path, not how long it takes to traverse it. Referring to Fig. 15(a), different paths that end at the same point in general give different phases, despite the fact that the time elapsed in traversing them is the same. Thus, $\gamma_n$ cannot be written as an explicit function of time. Referring to Figs. 15(b) and (c), using different amounts of time to traverse the same path yields the same phase. Said differently, for a given path, the phase accumulated in going from $R_0$ to $R$ is the same whether passage is carried out slowly or "less slowly," as long as it is done adiabatically. If $R$ returns to its initial value via a closed path $C$, the geometric phase $\gamma_n(C)$ is that of a completed circuit:

$$\gamma_n(C) = i \oint_C dR \cdot \langle n | \nabla n \rangle$$ (11.8)

Equation (11.8) reveals the gauge field. In Section 7 we saw that the coupling of a charged particle to the electromagnetic field yields the phase: $\int dr \cdot (q/c)A$, where $(q/c)A$ is a momentum. In the present case, the phase is $-\int dR \cdot \langle n | -i \nabla n \rangle$, so identification is clear: $\langle n | \nabla n \rangle$ (equivalently, $\langle n | \nabla n \rangle$ times a constant) is a gauge field.
If $|n\rangle$ is single valued its differentiation can be carried out with impunity along $C$. We will see that $|n\rangle$ can be assigned an arbitrary, parameter dependent phase without changing $\gamma_n(C)$, ensuring that a single valued wave function can be used. Even if $|n\rangle$ begins life not single valued, it can be made single valued, including the point at which the circuit closes. Alternatively, if the phase is chosen such that $d\mathbf{R} \cdot \langle n|\nabla n \rangle$ is zero along the path, it will be necessary to deal with a discontinuity upon completion of the circuit. This follows from the gauge invariance of $\gamma_n$.

Thus, there is a trade-off. The integrand $d\mathbf{R} \cdot \langle n|\nabla n \rangle$ can be made to vanish along $C$, but at the expense of a wave function that is not single valued. Alternatively, a single valued wave function can be used, but at the expense of nonzero $d\mathbf{R} \cdot \langle n|\nabla n \rangle$ along $C$. Of course the geometric phase $\gamma_n(C)$ is the same in either case.

In Section 10 (see Fig. 11) it was noted that the gauge field is a flux density in $\mathbf{R}$-space. The well known (i.e., found in most quantum mechanics textbooks) expression for probability current density: $j = (-i/2m)(\psi^* \nabla \psi - \psi^* \nabla \psi^*)$, provides intuition regarding eqn (11.8). We need to be careful with our use of the terms flux density and current density in the present context. The former passes through a surface enclosed by $C$, as indicated in Fig. 11. The latter flows on $C$. To underscore the correspondence between eqn (11.8) and the above expression for $j$, eqn (11.8) is written

$$\gamma_n(C) = i \frac{\hbar}{2} \oint_C d\mathbf{R} \cdot (\langle n|\nabla n \rangle - \{\nabla n|n \rangle \}) \tag{11.9}$$

In writing eqn (11.9), use has been made of the fact that $\langle n|\nabla n \rangle$ is imaginary because $\gamma_n(C)$ is real. Thus, eqn (11.8) can also be written

$$\gamma_n(C) = -\text{Im} \oint_C d\mathbf{R} \cdot \langle n|\nabla n \rangle \tag{11.10}$$

Geometric phase can be represented as a line integral of a current in parameter space, and through Stokes’ theorem as a flux density passing normal to the surface enclosed by $C$. Though eqns (11.8) and (11.10) are for a closed circuit, geometric phase accumulates along any path in $\mathbf{R}$-space. We shall focus on closed circuits, however, because manifestations of geometric phase are clear. The quantity $d\mathbf{R} \cdot \langle n|\nabla n \rangle$ is referred to as a connection because it relates the wave function to its increment: $d\mathbf{R} \cdot \langle \psi|\nabla \psi \rangle = \langle \psi|d\psi \rangle$.

This result is similar to what happens with classical vector transport on the surface of a sphere (Appendix 6), where the classical geometric phase is given by the enclosed solid angle. In the previous section we saw that quantum geometric phase is given by the enclosed solid angle in $\mathbf{R}$-space.

Choosing phase

We have seen that there are different ways to obtain $\gamma(C)$. For example, if the integrand vanishes along $C$, $\gamma(C)$ can be calculated at the close of the circuit where the wave

---

5 The fact that $\langle n|\nabla n \rangle$ is imaginary can also be obtained from normalization. Namely, $\langle n|n \rangle = 1$ on $C$, so its gradient vanishes: $0 = \{\nabla n|n \rangle + \langle n|\nabla n \rangle = \langle n|\nabla n \rangle + \langle n|\nabla n \rangle = 2 \text{Re} \langle n|\nabla n \rangle$.

240
function is discontinuous. The integral vanishing along $C$ means $\langle \psi | d\psi \rangle = 0$, i.e., $\psi$ is orthogonal to $d\psi$. This is not surprising. For example, in time independent non-degenerate time independent perturbation theory the first order correction to a wave function is always orthogonal to the wave function. An alternate way to obtain $\gamma(C)$ involves the use of a gauge transformation to make the wave function single valued on $C$.

These points are illustrated with an example in which the elements of a real $2 \times 2$ electronic Hamiltonian matrix depend parametrically on an angle in $R$-space.

$$ H = \begin{bmatrix} H_{11} & H_{12} \\ H_{12} & H_{22} \end{bmatrix} = \frac{1}{2} Tr H + R \begin{bmatrix} \cos \alpha & \sin \alpha \\ \sin \alpha & -\cos \alpha \end{bmatrix} \tag{11.11} $$

It is understood that the term $\frac{1}{2} \text{Tr} H$ is multiplied by a unit matrix.

The parameterization of the Hamiltonian that results in the angle $\alpha$ is given in the box. It can also be found in the quantum mechanics book of Atkins and Friedman, and I am sure other places as well. The eigenvectors $\psi_1$ and $\psi_2$ are those of the angle matrix. Assuming that the $H_{ij}$ matrix elements in eqn (11.11) are in a $\phi_1/\phi_2$ basis, the normalized eigenvectors are

$$ \begin{bmatrix} \psi_1 \\ \psi_2 \end{bmatrix} = \begin{bmatrix} \cos \alpha/2 & \sin \alpha/2 \\ -\sin \alpha/2 & \cos \alpha/2 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} \tag{11.12} $$

In electronic structure theory, the $\psi$'s are adiabats and the $\phi$'s are diabats. Equation (11.12) is a spin representation. Consequently, $\psi_{1,2}(2\pi) = -\psi_{1,2}(0)$. Also,

$$ \frac{\partial \psi_1(\alpha)}{\partial \alpha} = \frac{1}{2} \psi_2(\alpha) \quad \text{and} \quad \frac{\partial \psi_2(\alpha)}{\partial \alpha} = -\frac{1}{2} \psi_1(\alpha) \tag{11.13} $$

Note that $\langle \psi_1 | \nabla \psi_1 \rangle = \langle \psi_2 | \nabla \psi_2 \rangle = 0$ everywhere on $C$ except at $\alpha = 2\pi$, i.e., at the point where the wave function's phase is discontinuous. In this case, the geometric phase is calculated at the close of the circuit. For example,

$$ \gamma_1(C) = i \oint_C dR \cdot \langle \psi_1 | \nabla \psi_1 \rangle = i \int_{2\pi - \epsilon/2}^{2\pi + \epsilon/2} d\alpha \psi_1^\dagger (\partial \psi_1 / \partial \alpha) \tag{11.14} $$

241
At $\alpha = 2\pi$, $\psi_1$ undergoes a phase change of $\pi$. Thus, $\partial \psi_1 / \partial \alpha = \partial (\psi_1 e^{i\phi}) / \partial \alpha = i \psi_1 \partial \phi / \partial \alpha$, with $\phi$ accounting for the sign change of $\psi_1$. Setting $\partial \phi = \pi$ for small $\delta \alpha$ [i.e., $e$ in eqn (11.14)] yields $\gamma(C) = -\pi$.

Alternatively, the wave function is now gauge transformed. It is multiplied by $e^{i\alpha/2}$ in order to make it single valued.\footnote{To verify that the transformed wave function is single valued, multiply $\psi_1$ in eqn (11.12) by $e^{i\alpha/2}$ and write $\cos\alpha/2$ and $\sin\alpha/2$ in terms of exponentials:

$\tilde{\psi}_1 = e^{i\alpha/2} \left( \frac{1}{2} \left( e^{i\alpha/2} + e^{-i\alpha/2} \right) \phi_1 - i \frac{1}{2} \left( e^{i\alpha/2} - e^{-i\alpha/2} \right) \phi_2 \right) = \frac{1}{2} \left( (e^{i\alpha} + 1) \phi_1 - i(e^{i\alpha} - 1) \phi_2 \right)$, which is clearly single valued in $\alpha$.}

Using $\langle \tilde{\psi}_1 | \nabla \tilde{\psi}_1 \rangle = i \nabla \alpha / 2 = i \hat{e}_\alpha / 2$ in eqn (11.14) yields

$$
\gamma(C) = i \oint_C (d\alpha \hat{e}_\alpha) \cdot (i \hat{e}_\alpha / 2) = \frac{1}{2} \int_0^{2\pi} d\alpha = -\pi
$$

(11.15)

Thus, the equivalence of the approaches has been demonstrated. The geometric phase value of $-\pi$ (which could have been written down immediately from $\gamma(C) = \Omega/2$) has been obtained by integration on $C$ both with and without a single valued wave function. This illustrates that the gauge transformation affects how the calculation of the geometric phase is carried out, but not the result.

In the above, $e^{i\alpha/2}$ ensured single valuedness, and in so doing added what appears to be an angular momentum of $\frac{1}{2}$ to the nuclear dynamics. Suffice it to say that $\alpha$ depends on the Hamiltonian in a manner such that as $\alpha$ goes from 0 to $2\pi$ the system moves around a conical intersection. The $\frac{1}{2}$ is an intrinsic angular momentum that arises because of the $\psi_{1,2}$ degeneracy. It is present at energies well away from the degeneracy.

---

**Magnetic monopole: a mathematical curiosity**

If $A = \nabla \alpha \zeta = \hat{e}_\alpha / 2 \rho$ is a vector potential, its curl is a magnetic field. But $\nabla \times A$ is zero everywhere except the origin. To obtain $B$, use Stokes' theorem:

$$
\oint_C d\ell \cdot \nabla \zeta = \oint_C (\rho d\alpha \hat{e}_\alpha) \cdot (\hat{e}_\alpha / 2 \rho) = \pi = \iint_B \mathbf{S} \cdot B
$$

The fact that the circuit integral of $\mathbf{I} \cdot \nabla \zeta$ is $\pi$, whereas $B = \nabla \times \nabla \zeta$ vanishes everywhere except the origin means that $B$ is located at the origin. It is proportional to a delta function at $\mathbf{R} = 0 \ldots$ a monopole!
5.12. Summary

- The AB effect is a double slit experiment. If a measurement is carried out in which the passage of a charged particle causes the deflection of a device that contains $B$ or $E$, interference is eliminated because one of the slits in the double slit arrangement is blocked. Thus, the AB effect is a requirement of the quantum theory — without it quantum mechanics would not be consistent.
- The gauge field $A$ is a momentum that acts on the electron as an external parameter. In the AB effect, $A$ is not quantized. In Chapter 3, we saw how to quantize the transverse electromagnetic field (photons), whereas the $A$ encountered here is longitudinal (its curl vanishes, not its divergence). However, this $A$ can be quantized when it is the immediate consequence of quantized electron momentum. This is discussed in Appendix 5: Flux quantization in superconductivity.
- In the AB effect the electron passes slowly through the region where the gauge field is nonzero. Were this passage not slow, the electron might sense significant transient electric and magnetic fields. For example, in the magnetic AB effect, where the electron passes a cylinder that contains $B$, the electron would sense an electric field: $-\partial A / \partial t$. This can also be made clear by placing the electron in a box and transporting the box around the cylinder quickly. The electromagnetic fields have transverse components whose quanta are photons, so transitions between levels can take place, assuming the transient field has Fourier components whose frequencies match those of the electron in a box.
- In general terms, a gauge field is a momentum acting in a space of external parameters. When the external parameters are themselves the ingredients of a quantum theory, the momentum needs to be represented using a gradient. In the Born-Oppenheimer approximation, the external parameters are the nuclear coordinates, so the gauge field must be a gradient with respect to these coordinates.
Bibliography

These references are for Chapter 4 and Appendices 4 – 6.

1. Following eqn (5.6) is the statement: "... this result is valid only in regions where $B = \nabla \times A = 0."$ Prove that this is so.

2. Referring to Fig. 2, show that the only nonzero component of $A$ throughout the region $r > a$ is $A_\theta$.

3. In the electric version of the AB effect, the particle wave components are each sent through a hollow conducting tube as shown in the figure (under construction).

4. In obtaining eqn (3) of Appendix 5, a total derivative was replaced by a partial derivative. Explain why this is justified.

5. Consider the gauge transformation: $A \rightarrow A + \nabla \chi$ and $\psi \rightarrow \psi e^{iq\chi/\hbar}$. The position operator $r$ is invariant with respect to this transformation. Moreover, the canonical momentum operator $p = -i\hbar \nabla$ retains the same form it had before the transformation, thereby satisfying $[x,p] = -i\hbar$. Show that the expectation value of $p$ is not invariant with respect to the gauge transformation.

6. Show that $\pi$ is invariant with respect to: $A \rightarrow A + \nabla \zeta(r)$, $\psi \rightarrow \psi \exp(iq\zeta(r)/\hbar)$.

7. Consider the gauge transformation: $A \rightarrow A + \nabla \zeta(r)$ and $\psi \rightarrow \psi e^{iq\zeta(r)/\hbar}$. Note that $c = 1$ is used. The operator $\pi = p - qA - q\nabla \zeta(r)$ acting on the new wave function is the same as $\pi = p - qA$ acting on $\psi$. Why would one use the former? It has one more term than the latter and therefore involves more math. Is it not more complicated? Discuss this in terms of the Born-Oppenheimer approximation.

8. In this problem, you will consider a charged particle that is present in a time independent, spatially homogeneous magnetic field: how this appears in the Hamiltonian and some likely consequences of its presence.

   (a) Show that the vector potential $A$ is given by $-\frac{1}{2} \left( r \times B \right)$.

   (b) Assume $B = B_\hat{z}$. A charged particle is subject to a potential energy $V$ that is important overall but plays no role in the problem under consideration. In other words, we are only interested in effects due to $B$. Derive a Hamiltonian and put it in a compact form in terms of $H_0 = p^2/2m + V$ and $H'$, where $H'$ is due to $B$.

   (c) The term linear in $B$ is paramagnetism. It is responsible for what is called the normal (no spin) Zeeman effect. The term proportional to $B^2$ is diamagnetism. Assume that $B$ is small so its effects can be treated as perturbations. In the small-$B$ limit, the linear term dominates. Use a sensible basis and write the eigenvalues for this limit. Then add the quadratic term and explain how it can be treated. Comment on the result.

9. Following eqn (10.18) is the statement: "Some algebra yields $V_m = (m/B^2)\hat{e}_y."$ Fill in these steps.
Aharonov-Bohm effect and geometric phase

10. Consider the Zeeman effect for a one-electron-type atom having \( l = 1 \) and \( s = \frac{1}{2} \), e.g., atomic hydrogen with \( n = 2 \). The Zeeman Hamiltonian is \( H_Z = A(l + 2s) \cdot B \), where \( A \) is a negative real constant and \( B \) is the external magnetic field. Calculate the geometric phase for a closed circuit that encloses the \( B = 0 \) degeneracy point. Do this for the \(|1, +\rangle\) and \(|0, +\rangle\) states, where 1 and 0 denote \( m_l \) values, and + denotes \( m_s = +\frac{1}{2} \). The geometric phases for the other \( H_Z \) eigenstates can be inferred. Results were obtained in the text for integer and odd-half-integer spin, where the gradient of the Hamiltonian is proportional to the spin vector. Here, \( H_Z = A(l + 2s) \cdot B \) is examined to see if the interesting factor of two has an effect, and if so, what is it.

11. The \( n = 2 \) state of atomic hydrogen has an eightfold degeneracy, not counting nuclear spin. Neglect nuclear spin. That is, \( \Sigma (2l + 1)(2s + 1) \) for \( s = 1/2 \) and \( l = 0 \) and 1 is equal to 8. The summation is over \( l \) values. An H atom is present in a steady electric field \( E = E_0 \vec{e}_z \). To begin, neglect spin-orbit interaction. The Stark Hamiltonian is \( H_S = -\mu \cdot E = q r \cdot E \), where \( \mu = qr = -e r \) is the electric dipole moment operator. Thus, \( H_S = E_0 r \cos \theta \). When \( E = 0 \), the \( n = 2 \) levels are degenerate, but when \( E \neq 0 \), this degeneracy is lifted for \( 2p_z \) and \( 2s \), which mix, creating a large dipole moment. This problem should be treated as a project that might require several days.

(a) The magnitude of \( E \) remains constant, but \( E \) now slowly changes its orientation relative to the laboratory reference frame. Its tip traces a closed circuit \( C \) whose surface contains the \( E = 0 \) degeneracy point. Make a picture that shows how the dipole moment follows \( E \). Discuss this system in the context of geometric phase. How much geometric phase has accrued upon completion of the circuit? Is it possible to deduce the answer without resorting to a lengthy calculation? If so, how, what is the answer?

(b) Now introduce spin-orbit interaction. Examine the limits of large and small \( E_0 \) and calculate the energies. In these respective limits, the Stark effect and spin-orbit interaction dominate, and the weaker of the two interactions can be treated perturbatively. Again, discuss the system within the context of geometric phase.

(c) Discuss qualitatively the case \( E \neq 0 \) and \( B \neq 0 \), where \( B \) is a magnetic field; ignore spin-orbit interaction. Each field can be varied independently, both magnitude and orientation. What are the external parameters? What is the dimension of the \( R \)-space?

12. In Appendix 6, it was stated in the box entitled Coriolis force that \( x(t) \) and \( y(t) \) have a phase difference of \( \pi/2 \). Show that this is true.

13. An astronaut goes to Mars and brings along a Foucault pendulum. Once on the planet, the astronaut measures the precession angle and finds that in 24 hours the pendulum has precessed 180° in the clockwise direction. Where is the pendulum located relative to the Martian equator and north pole?

14. Consider vector transport on the surface of a sphere (Appendix 6). Using a model sphere, review the derivation of the Gauss-Bonnet formula and the holonomy of a triangle. Extend this to a closed four-sided path, which obviously can be made from two triangles joined by a common side. Now extend it to a regular polygon of arbitrary shape.
15. (Courtesy of Phil Pechukas) The adiabatic theorem has to do with motion under a
time dependent Hamiltonian, and it says (roughly) once in an eigenstate, always in an
eigenstate, provided the Hamiltonian changes sufficiently slowly. Let $H(\lambda), 0 \leq \lambda \leq 1,$ be a Hamiltonian that depends smoothly on a parameter $\lambda.$ The adiabatic theorem considers the problem

$$i\hbar \dot{\psi}_\epsilon(t) = H_\epsilon(t) \psi_\epsilon(t)$$  \hspace{1cm} (i)

where $H_\epsilon(t) = H(\epsilon t),$ and then looks at the limit $\epsilon \to 0.$ Suppose instead we set $\epsilon = 1$ and vary $\hbar,$ looking at the problem:

$$i\hbar \dot{\phi}_\hbar(t) = H(t) \phi_\hbar(t)$$  \hspace{1cm} (ii)

as $\hbar \to 0.$ Let $\hbar_0$ denote the real Planck’s constant. Show that $\phi_\hbar(t) = \psi_{\hbar/\hbar_0}(\hbar_0 t/\hbar).$ Thus, the adiabatic limit is equivalent to the quasiclassical limit $\hbar \to 0.$

16. (Courtesy of Phil Pechukas) A magnetic field $B(t)$ lies in the $x$-$y$ plane and slowly
precesses around the $z$-axis: $B_z(t) = 0; B_x(t) = B \cos \alpha t; B_y(t) = B \sin \alpha t.$ The Hamiltonian for a spin-$1/2$ particle in the field looks like this:

$$H(t) = \frac{\Delta}{2} \begin{pmatrix} 0 & e^{-i\alpha t} \\ e^{i\alpha t} & 0 \end{pmatrix}$$  \hspace{1cm} (i)

where the energy splitting $\Delta$ is proportional to the field strength $B.$ Solve the Schrödinger equation

$$i\hbar \left( \begin{array}{c} \dot{c}_1(t) \\ \dot{c}_2(t) \end{array} \right) = H(t) \left( \begin{array}{c} c_1(t) \\ c_2(t) \end{array} \right)$$  \hspace{1cm} (ii)

exactly for the case $c_1(0) = c_2(0) = 1;$ using 1 instead of $1/\sqrt{2}$ simplifies manipulations and is fixed trivially later. Hint: define $\tilde{c}$ by writing

$$\left( \begin{array}{c} c_1(t) \\ c_2(t) \end{array} \right) = \left( e^{-i\alpha t} 0 \\ 0 e^{i\alpha t} \right) \left( \begin{array}{c} \tilde{c}_1(t) \\ \tilde{c}_2(t) \end{array} \right)$$  \hspace{1cm} (iii)

By a clever choice of $\alpha$ you will get an equation for

$$\left( \begin{array}{c} \dot{\tilde{c}}_1(t) \\ \dot{\tilde{c}}_2(t) \end{array} \right)$$  \hspace{1cm} (iv)

that you know how to solve. Show that the exact solution stays close to the $+ \Delta/2$
eigenvector of $H(t)$ for all time, provided $\omega$ is sufficiently small. What is the condition that $\omega$ must satisfy to be considered sufficiently small?