The geometric phase

Jeeva Anandan

When a quantum system evolves so that it returns to its initial physical state, it acquires a 'memory' of this motion in the form of a geometric phase in the wavefunction. This phase has observable consequences in a wide range of physical systems, and its presence has now been convincingly demonstrated, for example, in optical and nuclear magnetic resonance experiments.

A THEORY of nature, once well understood, may not be expected to have surprises in store for us. It is therefore a testament to the richness of quantum theory that it continues to surprise us many decades after its initial formulation. One such surprise was the discovery in 1959 by Aharonov and Bohm[1] that the behaviour of an electrically charged quantum system can be altered in an apparently non-local manner—that is, without any forces acting on it. If a beam of electrons is split into two and the two beams are made to interfere by passing around a cylinder (Fig. 1), interference fringes are observed in the form of a variation in the intensity of the combined beam. Aharonov and Bohm predicted that if a magnetic field is introduced inside the cylinder, these interference fringes will shift even though no electric or magnetic field lines intersect the beams (so that no forces act on them).

Another surprise was the discovery by Berry[2] that a slowly evolving (adiabatic) quantum system retains a 'memory' of its motion when returned to its original physical state. This 'memory' is termed Berry's phase or the geometric phase. The appearance of a geometric phase has been generalized to the case of non-adiabatic evolution[3] of quantum systems, which helps to underline the purely geometrical nature of the phenomenon (which is to say, it depends purely on the geometry of the pathway along which the system evolves). There have now been several experimental observations of the adiabatic and non-adiabatic geometric phases, and it has become clear that it is related to a wide range of physical phenomena, both in quantum systems (for which the Aharonov–Bohm effect represents one expression) and in classical systems (where the corresponding effect is exemplified by the motion of the Foucault pendulum). Here I shall describe how the geometric phase arises in quantum systems; this is seen most easily by first considering its classical analogue. I shall also review briefly some of the ways in which the geometric phase has been observed experimentally—in optics, neutron optics and nuclear magnetic resonance studies.

The role of phase

According to quantum theory, the state of a particle at any given time is described by a wavefunction ψ, which is a complex-valued function of space. As the system evolves in time this wavefunction changes accordingly. In the evolution considered by Berry[2], the final wavefunction ψ' is related to the initial wavefunction ψ by $\psi' = e^{i\phi}\psi$, where ϕ is a real number and $i^2 = -1$. This means that the observable characteristics of the initial and final states are the same and therefore according to classical physics we cannot distinguish between them. But in quantum mechanics, ϕ, which is the additional phase acquired by the wavefunction, constitutes a 'memory' of the evolution undergone by the system. Berry showed that, in addition to a contribution from the effect of the environment at each instant of time, ϕ has a part that is geometrical in nature.

As a simple example, consider a neutron rotating in a magnetic field. A neutron has spin angular momentum $S = \frac{1}{2}\hbar$, where $\hbar = h/2\pi$ and h is Planck's constant. With any given state of the spin we can associate a unique direction which may be visualized as the average direction of the spin. We may therefore define a spin vector **S** pointing in this direction with magnitude S. If the neutron is rotated about this spin direction by an angle θ, its wavefunction acquires a phase $\phi = -\frac{1}{2}\theta$. For example, in a magnetic field **B**, the interaction between **B** and the neutron's magnetic moment $\boldsymbol{\mu}$ (which lies along the spin direction and points in the opposite direction) would rotate the neutron continuously about **B**; this is called the Larmor precession. Therefore if **B** is in the direction of the neutron spin, the phase of **S** will change continuously.

Suppose now that the direction of **B** is changed slowly so as to return eventually to its original direction. The neutron spin direction will be pinned to **B** and will move with it slowly (that is, adiabatically) (Fig. 2*a*). But the total rotation of the neutron about its spin direction is not, in general, equal to just the accumulated rotation owing to its Larmor precession at every instant: there is an additional rotation α, which results in an additional change of phase of the wavefunction $\beta = -\frac{1}{2}\alpha$. This β is Berry's phase.

Berry's approach was to use a set of parameters that define the environment of the quantum system. In the above example, the magnetic field is the environment and these parameters are the components of the magnetic field. The additional rotation α is equal to the solid angle subtended at the origin by the curve Γ traced out by **B** in this three-dimensional parameter space. This result has been verified experimentally; I shall derive it below and show that it is purely geometric in origin.

Berry's phase was reconsidered by Aharonov and Anandan[3,4], who shifted the emphasis from changes in the environment to the motion of the quantum system. In the above example, for instance, they associated the geometric phase with the motion of the neutron spin and not the motion of the magnetic field. This is useful because there are an infinite number of ways of varying the magnetic field to obtain a given motion of the spin, and Aharonov and Anandan showed that, for all of them, the

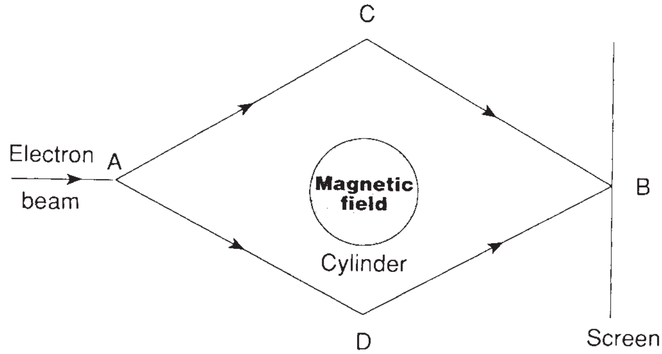

FIG. 1 Schematic illustration of the Aharonov–Bohm effect. An electron beam is split into two at A and the two resulting beams are diverted at C and D so that they interfere in the region surrounding B; the interference pattern is observed on the screen. A magnetic field confined to the cylinder shifts the phase difference between the interfering beams. This results in a shift in the interference pattern, even though the beams move in a field-free region with no forces acting on them.

same geometric phase is obtained which is uniquely associated with the motion of the neutron. This approach enabled them to generalize Berry's phase to non-adiabatic motions. For example, when the magnetic field is changed rapidly (regardless of whether it returns to its original direction), if the neutron spin returns to its original state then it will have undergone an additional rotation equal to the solid angle α subtended by the curve C traced out by the spin vector S (Fig. 2*b*). This gives rise to a geometric phase $\beta = -\frac{1}{2}\alpha$; but now it is not necessary for the motion to be adiabatic. Note that the neutron also has another state with spin vector −S which, under the same circumstances, changes by a phase that includes a geometric phase $\beta = +\frac{1}{2}\alpha$.

The approach of Aharonov and Anandan is in the spirit of the work of Pancharatnam[5], who introduced a geometric phase in optics about three decades before Berry, in his study of the rotation of the polarization of light owing to a sequence of measurements. The polarization state of a photon is analogous to the spin of the neutron. It turns out that the Aharonov–Bohm effect and the geometric phase can both be explained by basically the same geometric concept, called parallel transport.

Classical parallel transport

For simplicity, consider first a vector on a two-dimensional plane which is parallel-transported along a curve: that is, moved along the curve so that its length and direction are constant. Parallel transport may be generalized to a vector moved along a curve on a curved surface such as a sphere by requiring that the vector moves without a change in its magnitude and without rotation about the instantaneous normal to the surface; that is, the component of the angular velocity of the vector in the direction of the normal is zero. This condition may be restated in terms of concepts that are intrinsic to the surface, by noting that a small portion of a smooth surface is like a plane. Therefore parallel transport may be defined along any smooth curve by requiring that, for each infinitesimal segment of the curve, it is the same as parallel transport on a plane: the given vector, the infinitesimal segment of the curve and the parallel-transported vector form three sides of a parallelogram. This restatement enables us to ignore the three-dimensional space in which the surface is embedded. (If the curve is in a curved space of arbitrary dimension, instead of on a two-dimensional surface, parallel transport may be defined in the same way.) In general, a suitable rule or prescription for parallel transport of a vector along any curve in an arbitrary space is called a connection. The special case of the infinitesimal parallelogram rule above is called a Riemannian connection.

Of great importance is the transformation undergone by vectors when they are parallel-transported around a closed curve, which is called a holonomy transformation. If a vector is parallel-transported around a closed curve on the plane it comes back unchanged, but if parallel-transported around a closed curve C on a sphere it gets rotated by an angle equal to the solid angle subtended by C at the sphere's centre, which is the holonomy transformation in this case (Fig. 3*a*). For a sphere of radius r, this solid angle is given by

$$\alpha = A/r^2 \qquad (1)$$

where A is the area enclosed by the closed curve, and the curvature is $1/r^2$ everywhere on the sphere. This result can be seen quite easily for the case of a vector, initially at the north pole, which is parallel-transported first to the equator by moving it down the longitude along which the vector points, then a quarter of the way around the equator, and finally back to the north pole along the new longitude. The vector ends up pointing along the new longitude (Fig. 3*b*). Therefore, even though the vector has not been rotated at any point along its route, nevertheless it arrives back at the north pole rotated by the angle between the two longitudes, namely $\pi/2$ radians. This is also the solid angle subtended at the centre of the sphere by the vector's path, as obtained from equation (1).

Parallel transport on a sphere can be illustrated by means of the Foucault pendulum, which is used to measure the rate of rotation of the Earth. Imagine a set (triad) of cartesian axes on the surface of the Earth (approximated as a sphere) with its z axis always vertical and the x and y axes always horizontal. A Foucault pendulum, suspended from a point on the z axis, oscillates in a vertical plane. If the triad is moved so that its x and y axes are parallel-transported along a curve C on the Earth's surface, the triad acquires no rotational velocity about the z axis (normal to the surface of the Earth) and therefore the plane of oscillation of the pendulum with respect to this triad does not change. Conversely, we can use the condition that the plane of oscillation of the Foucault pendulum is constant to ensure that the x and y axes are being parallel-transported. When the triad returns to its original position, the plane of oscillation has rotated by an angle α, equal to the solid angle subtended by C at the centre of the Earth[8]. This is an example of Hannay's angle[9] which is a classical analogue of Berry's phase.

This experiment has so far been performed keeping the point of oscillation fixed with respect to the Earth, so that as the Earth rotates, C runs around a latitude of the Earth. Then $\alpha = 2\pi(1-\cos\theta)$ or $2\pi\cos\theta$, depending on the sense in which it is measured (Fig. 3*c*). But clearly, by moving the point of oscillation appropriately, C can be any closed curve on the Earth's surface. Thus the Foucault pendulum, which is usually used to demonstrate that the Earth rotates, can just as well be used to demonstrate, in principle, that the Earth's surface is curved. If the Earth were a disk then α would always be 0 (or 2π), corresponding to parallel transport on a plane. The Earth's radius r may be determined from equation (1) if we know α and the area of the Earth's surface enclosed by C.

In practice such a determination could be made by having a Foucault pendulum on each of two nearby ships with their planes of oscillations parallel. The ships now separate for a while and then come together again so that the closed curve formed by their paths on an imaginary non-rotating sphere that coincides with the Earth's surface enclosed an area A. The planes of oscillation will no longer be parallel, but will diverge at an

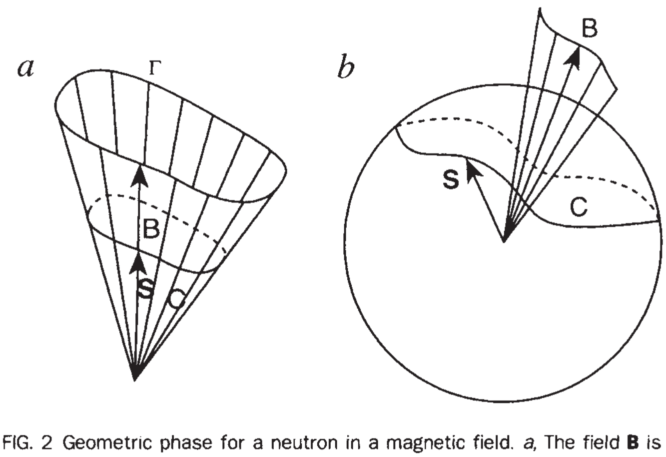

FIG. 2 Geometric phase for a neutron in a magnetic field. *a*, The field **B** is varied slowly along the curve Γ. The neutron spin, represented by the vector **S**, follows **B**. The phase acquired by the neutron is the sum of the dynamical phase that it acquires through its Larmor precession in **B**, and the geometric phase, equal in magnitude to half the solid angle α subtended by Γ at the origin. *b*, **B** moves in such a way that the neutron spin vector **S** traces the circuit C on a sphere with radius *S*. Unlike case *a*, the magnetic field need not be parallel to **S** or vary slowly. The phase acquired is the sum of the dynamical phase resulting from the Larmor precession around the component of **B** in the direction of **S** at every instant of time, and the geometric phase whose magnitude is half the solid angle subtended by C at the origin. *a*, is obtained as a special case corresponding to a particular choice of **B** that is parallel to **S** and slowly varying.

angle α, and r is then determined from equation (1). This approach is fundamentally different from that of the ancient Greeks, who measured r by comparing the shadows of vertical poles at two different points on the Earth's surface, because the latter is a measurement of extrinsic curvature—the curvature of the Earth in the three-dimensional space in which it is embedded—whereas the former is a measurement of intrinsic curvature, defined by intrinsic properties of the surface without any reference to the euclidean geometry of the embedding space.

But the Earth is strictly not a sphere: it is slightly oblate. For an arbitrary surface, the intrinsic curvature R (sometimes called the gaussian curvature) at any point may be defined as follows: the angle by which a vector rotates when it is parallel-transported around an infinitesimal closed curve around this point is

$$d\alpha = R\, dA$$

where dA is the area of the infinitesimal portion enclosed by the curve. For a curve that encloses a finite area, the area can be divided into infinitesimal areas and the angle of rotation α of a vector parallel-transported around this curve (the holonomy transformation) is the sum of the contributions from the infinitesimal areas:

$$\alpha = \int R\, dA \qquad (2)$$

In general R can vary from point to point, but for a sphere R is equal to $1/r^2$ everywhere. Therefore equation (2) gives equation (1) in this special case.

Geometric phases and parallel transport

Classical parallel transport should enable us at least to gain some intuitive understanding of the Aharonov–Bohm effect and Berry's phase. The former is analogous to parallel transport on a cone. A cone may be formed by taking a flat piece of paper bounded by two straight edges at an angle θ and joining these edges (Fig. 4). As the paper is not stretched or compressed during this process, the cone has no intrinsic curvature except at the vertex, which may be smoothed out so that the curvature is finite everywhere. Thus, except for this region at the vertex, the intrinsic geometry in any small region of the cone is the same as the flat geometry of the initial sheet of paper. Therefore, away from the vertex, a vector will be parallel-transported just as it would be on the flat sheet of paper. But a vector **V** parallel-transported into **V**′ around a closed curve C that encloses the vertex undergoes a rotation by the angle θ. This is the holonomy transformation associated with C.

The curvature at the vertex can be regarded as analogous to the magnetic field within the cylinder in the Aharonov–Bohm experiment, while the zero intrinsic curvature everywhere else corresponds to the vanishing of the magnetic field outside the cylinder. Then the rotation by the angle θ which relates **V** and **V**′ (Fig. 4) is analogous to the phase difference between the two interfering beams, which they acquire by travelling in the field-free region, just as **V** moves in the curvature-free region. The fact that the Aharonov–Bohm effect cannot be understood in terms of forces suggests that this analogy with parallel transport should be taken seriously.

This analogy is valid more generally when both electric and magnetic fields are present. This suggests that the electromagnetic field may be a connection for parallel transport of the value of the wavefunction $\psi(\mathbf{x}, t)$ along a path in space–time, and the Aharonov–Bohm effect arises from the corresponding holonomy transformation around the closed curve followed by the interfering beams. The statement that the electromagnetic field is such a connection is the same as saying that it is a gauge field. The electric and magnetic field strengths at each point in space-time constitute the curvature of this connection at that point, with the connection represented by the electromagnetic potential.

Classical parallel transport can also help us to understand Berry's phase for the neutron in a magnetic field, mentioned previously (Fig. 2*a*), as pointed out by Anandan and Stodolsky[6]. As the direction of the magnetic field **B** changes slowly, the spin vector **S** (which follows **B**) will trace out a curve C on a sphere with radius *S*. Imagine now a cartesian triad moved around C with its z-axis radial and its x- and y-axes being parallel-transported. It returns rotated about the z axis by the solid angle α given by equation (1). Similarly, a neutron spin vector that maintains its orientation relative to the triad will also be rotated by the same angle α. This rotation, the origin of which is purely geometrical, changes the phase of the spin wavefunction by $\beta = -\frac{1}{2}\alpha$ because the axis of rotation here is along the direction of the spin vector. At each instant, the neutron also undergoes a Larmor precession with frequency $\omega_L = 2\mu B/\hbar$, where μ is the magnitude of its magnetic moment $\boldsymbol{\mu}$. The accumulation of these rotations leads to an additional phase change, called the dynamical phase, owing to the dynamical interaction between $\boldsymbol{\mu}$ and **B**.

Even this simple example shows the magnitude of Berry's achievement: he effectively identified a new way of rotating the neutron spin which had been missed by all the physicists who for many years had studied the properties of the neutron. This rotation is purely geometrical, so that, unlike the Larmor precession, it is not necessary to know the values of the dynamical quantities such as μ, B or even h in order to determine it[6].

But by focusing on the motion of **S** instead of **B** we can see that the two restrictions on **B** imposed by Berry—adiabaticity and a need for **B** to be parallel or antiparallel to **S**—may be removed. Consider the more general case in which **S** traces out the same curve C but **B** need be neither slowly varying nor in the direction of the neutron spin (Fig. 2*b*). There is an infinite number of ways of changing $\mathbf{B}(t)$ as a function of time t to make this happen. At any instant t there is a Larmor precession of the neutron with frequency ω_L about the direction of $\mathbf{B}(t)$. Hence the angular velocity of the neutron about its spin direction is equal to the projection of the total angular velocity onto the spin direction. Therefore the angular frequency of the precession about the spin axis with respect to the triad

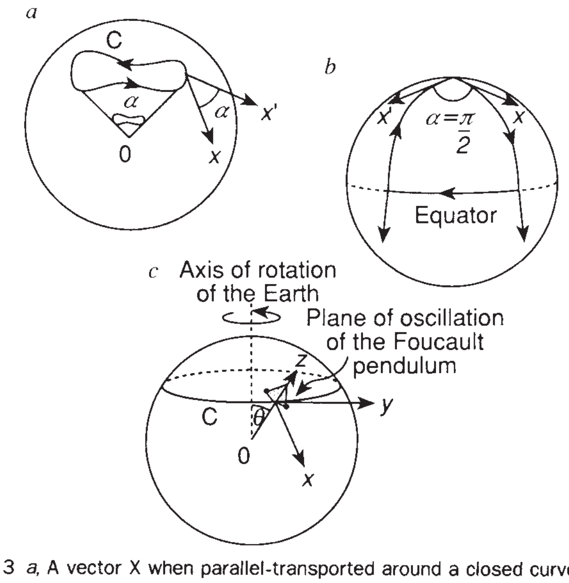

FIG. 3 *a*, A vector X when parallel-transported around a closed curve C on a sphere returns as X′, which is related to X by a rotation by an angle equal to the solid angle subtended by C at the centre of the sphere. *b*, Verification of this result for the special case of parallel transport down a longitude, around the equator by $\pi/2$, and back up the new longitude. In this case $\alpha = \pi/2$. *c*, Parallel transport illustrated by the Foucault pendulum. The imaginary cartesian triad moves along the latitude C so that its z-axis is radial while the x- and the y-axis are parallel-transported on the Earth's surface. The plane of oscillation of the pendulum is constant with respect to this triad. But it comes back rotated by the same angle α by which the triad has rotated about the z-axis because of parallel transport around C.

parallel-transported along C is $\omega_L \cos\gamma = (2\mu B\cos\gamma)/\hbar$, where γ is the angle between **B** and **S**. The accumulated rotation owing to this precession gives the dynamical phase. This generalizes the dynamical phase of Berry, which corresponds to the special case $\gamma = 0$. But as we have seen, the triad parallel-transported around C comes back rotated by the solid angle α, so that the neutron's spin wavefunction acquires an additional, geometric phase $\beta = -\frac{1}{2}\alpha$. The geometric phase depends only on the motion of the system, specified by the curve C, and is independent of the orientation or rate of change of the magnetic field that causes this motion.

Quantum parallel transport

Thus we can understand the geometric phase for spin by considering classical parallel transport on a sphere and its effect on the wavefunction. For more-complex quantum systems this argument can be generalized using group theory, in which case α is generalized to a set of geometric angles[6,7], described briefly later. But the origin of the geometric phase in an arbitrary quantum system can also be understood by considering parallel transport within the context of quantum theory, in which the state of a system is described by a vector.

There are two important fundamental differences between classical and quantum physics. First, in quantum physics one may form a new state of a system by adding two initial states (which is why quantum states are represented by vectors), whereas in classical physics adding dynamical states is not possible. More generally, we can form any linear combination of two quantum states (see Box 1). For example, for electron beams interfering around a cylinder (Fig. 1) the intensity distribution in the interference region can be explained by supposing that the state of each electron in this region is the sum of two states corresponding to the two beams, as if the electron traveled simultaneously along both beams. But according to classical physics, the electron can travel along only one or other of the beams, giving rise to a different intensity distribution that does not accord with experiment.

Another important difference is that, if a quantum system is known to be in a state described by the vector $|\psi_1\rangle$ and a measurement is performed to see if it is in a state with vector $|\psi_2\rangle$, quantum theory predicts a probability for the success of this experiment, called the probability of transition. In classical physics, on the other hand, a measurement made on a system in a given state will not find it in another state. The quantum probability of transition is evaluated from the 'probability amplitude' for a transition from state $|\psi_1\rangle$ to $|\psi_2\rangle$, a complex number that depends only on the two state vectors. The probability amplitude is written $\langle\psi_2|\psi_1\rangle$, and is also called the inner product between $|\psi_1\rangle$ and $|\psi_2\rangle$ (see Box 1). We can write $\langle\psi_1|\psi_2\rangle = A\,e^{i\phi}$, where ϕ is a real number representing the phase difference between $|\psi_1\rangle$ and $|\psi_2\rangle$, and A is a non-negative real number not exceeding unity. Then $|\langle\psi_1|\psi_2\rangle|^2 = A^2$ is the probability of transition from $|\psi_1\rangle$ to $|\psi_2\rangle$.

Thus A corresponds to an observable quantity. But does ϕ, which may seem to arise from a mathematical formalism, also have observable physical consequences? We shall see that the geometric phase is a direct physical consequence of this phase difference. This aspect of the probability amplitude was perhaps first realized by Pancharatnam[5], who showed that if two state vectors $|\psi_1\rangle$ and $|\psi_2\rangle$ have the same phase (that is, $\phi = 0$), and $|\psi_2\rangle$ and $|\psi_3\rangle$ also have the same phase, then $|\psi_1\rangle$ and $|\psi_3\rangle$ need not have the same phase. The non-transitivity of the relation of two state vectors having the same phase can be understood in terms of parallel transport, and it is this non-transitivity that is ultimately responsible for the geometric phase[4,12-14].

To determine the physical state of a quantum system, we need to measure a set of compatible observables. If the state vector $|\psi\rangle$ gives a certain set of observed values for these observables, then $|\psi'\rangle = c|\psi\rangle$, where c is any complex number, will give the same values (see box). The set of all vectors obtained by multiplying a given vector $|\psi\rangle$ by all possible complex numbers is called a ray. It follows that all vectors in a given ray represent the same physical state because they give the same values for the measured observables.

The set of rays of the Hilbert space $\mathcal{H}$ (see Box 1) is called the projective Hilbert space[3], denoted by $\mathcal{P}$. For neutron spin states, for example, $\mathcal{P}$ can be identified with the set of spin vectors **S** which form the sphere that I used earlier to explain the geometric phase for the neutron. Consider a curve p in $\mathcal{P}$, such that $p(s)$ denotes the point on this curve corresponding to the parameter value s. Also, for each value of s let $|\psi(s)\rangle$ be a vector in the ray corresponding to $p(s)$ such that $|\psi(s)\rangle$ varies smoothly along the curve $p(s)$. Two natural conditions for parallel-transporting $|\psi(s)\rangle$ along a curve are that the length of $|\psi(s)\rangle$ is preserved (that is, $\langle\psi(s)|\psi(s)\rangle$ is independent of s) and that $|\psi(s)\rangle$ and $|\psi(s+ds)\rangle$ have the same phase (that is, $\langle\psi(s)|\psi(s+ds)\rangle$ is real and positive). These two conditions imply that

$$\langle\psi(s)|\frac{d}{ds}|\psi(s)\rangle = 0 \tag{3}$$

which was obtained in a different manner by Simon[15].

Equation (3) is the rule for parallel transport of $|\psi(s)\rangle$ along any curve in $\mathcal{P}$, and therefore defines a connection over $\mathcal{P}$. It indicates that the change in $|\psi(s)\rangle$ during parallel transport, $d|\psi(s)\rangle/ds$, has a zero inner product with $|\psi(s)\rangle$, so that the change in $|\psi(s)\rangle$ cannot be along its own direction. As $\langle\psi(s)|\psi(s)\rangle$ is independent of s, the only way $|\psi(s)\rangle$ can change along its own direction is by being multiplied by a phase factor $e^{i\varepsilon}$, which may be regarded as a 'rotation' of $|\psi(s)\rangle$. Hence, forbidding such a 'rotation' is analogous to the no-rotation condition in the definition of parallel transport of a vector on a sphere described earlier. If, however, $|\psi(s)\rangle$ is parallel-transported around a closed curve ($p(\tau) = p(0)$ for some τ), it comes back 'rotated':

$$|\psi(\tau)\rangle = e^{i\beta}|\psi(0)\rangle \tag{4}$$

BOX 1 Some fundamental principles of quantum mechanics

According to the notation introduced by Dirac, a state (vector) of a quantum system is denoted by the 'ket' $|\psi\rangle$. Given two states $|\psi_1\rangle$ and $|\psi_2\rangle$, any linear combination $c_1|\psi_1\rangle + c_2|\psi_2\rangle$, where c_1 and c_2 are complex numbers, represents a possible new state. If $\mathcal{H}$ denotes the set of possible states of a quantum system then the above property is contained in the statement that $\mathcal{H}$ forms a vector space over the field of complex numbers. Given two states $|\psi\rangle$ and $e^{i\theta}|\psi\rangle$, where θ is real and $i^2 = -1$, we say that θ is the phase difference between the two states.

The inner product of two given states $|\psi_1\rangle$ and $|\psi_2\rangle$ is a complex number, denoted $\langle\psi_1|\psi_2\rangle$. The vector space $\mathcal{H}$ with a suitable inner product is called a Hilbert space. An observable is a measurable property of a quantum system and is represented by an hermitian operator that acts on the Hilbert space $\mathcal{H}$. During a measurement of an observable, the possible values obtained, which are real, are the eigenvalues of the corresponding hermitian operator. (The eigenvalues of an hermitian operator are always real.)

The dynamical evolution of a state is given by the Schrödinger equation.

$$i\hbar\frac{d}{dt}|\psi\rangle = H|\psi\rangle$$

where H is a linear hermitian operator acting on the states in $\mathcal{H}$, called the hamiltonian, and $\hbar = h/2\pi$, h being Planck's constant. It follows that the Schrödinger equation is linear in the sense that, if $|\psi_1(t)\rangle$ and $|\psi_2(t)\rangle$ are two solutions of this equation, the complex linear combination $c_1|\psi_1(t)\rangle + c_2|\psi_2(t)\rangle$, where c_1 and c_2 are arbitrary complex numbers, is also a solution.

The wavefunction of a quantum state is defined by $\psi(\mathbf{r}, t) = \langle\mathbf{r}|\psi(t)\rangle$, where $|\mathbf{r}\rangle$ is the state corresponding to the system being in position $\mathbf{r}$. Then $|\psi(\mathbf{r}, t)|^2$ is the probability per unit volume of finding the system at position $\mathbf{r}$ at time t.

Thus the factor $e^{i\beta}$ arises from the holonomy transformation in this case, and β is the geometric phase. This additional phase is acquired because of the curvature of the connection on $\mathscr{P}$. I will now derive an explicit expression for β.

Calculating the geometric phase

We say that an evolution $|\psi(t)\rangle$ is cyclic in the time interval $t=0$ to τ if $|\psi(\tau)\rangle = e^{i\phi}|\psi(0)\rangle$ for some phase ϕ. As points in $\mathscr{P}$ have no phase information, the beginning and end points of this evolution must correspond to the same point in $\mathscr{P}$, so that the system moves around a closed curve in $\mathscr{P}$. We now define a basis vector field $|\tilde{\psi}(t)\rangle$ such that $|\psi(t)\rangle = e^{if(t)}|\tilde{\psi}(t)\rangle$ and $|\tilde{\psi}(\tau)\rangle = |\tilde{\psi}(0)\rangle$. In this way we express the phase of $|\psi(t)\rangle$ in terms of $|\tilde{\psi}(t)\rangle$. This is analogous to describing the vector on a curved surface by its components with respect to a chosen coordinate system on the surface. The parallel transport of the vector can be described in terms of these components. Part of the phase change ϕ is due to the quantum parallel transport of $|\psi(t)\rangle$, which can be expressed in terms $|\tilde{\psi}(t)\rangle$. This phase, denoted β, generalizes to an arbitrary quantum system the geometric phase obtained earlier for the neutron. It can be shown[3], when $|\psi(t)\rangle$ evolves according to the Schrödinger equation (see box), that $\phi = \beta + \delta$, where

$$\delta = -\hbar^{-1}\int_0^\tau \langle\psi(t)|H(t)|\psi(t)\rangle\,\mathrm{d}t \tag{5}$$

is the dynamical phase and

$$\beta = i\int_0^\tau \langle\tilde{\psi}(t)|\frac{\mathrm{d}}{\mathrm{d}t}|\tilde{\psi}(t)\rangle\,\mathrm{d}t = i\oint_C \langle\tilde{\psi}|\,\mathrm{d}\,|\tilde{\psi}\rangle \tag{6}$$

is the geometric phase; here d is the differential operator on $\mathscr{P}$ and $H(t)$ is the Hamiltonian operator, the eigenvalues of which are the energy levels of the system.

The phase β is geometrical because of the following reasons: (1) Unlike δ, β is independent of the rate at which the system travels along the unparametrized curve C. (2) Unlike δ, β is independent of the particular hamiltonian $H(t)$ producing motion along C. There are, in fact, an infinite number of hamiltonians that cause a given motion of the quantum systems along C, but the same geometric phase β is obtained for all of them. (3) The geometric phase factor $e^{i\beta}$ is independent[4] of the chosen $|\tilde{\psi}\rangle$. Hence the factor $e^{i\beta}$ depends only on C—it corresponds to the holonomy transformation associated with C for the connection defined by equation (3). In other words, if $|\psi\rangle$ is parallel-transported around C, it gets multiplied by $e^{i\beta}$. Statements (1), (2) and (3) show that β is geometrical in origin because it depends only on the motion of the quantum system defined by C and not on the particular interaction undergone by the system, which is determined by H.

For the simple case of the neutron considered earlier, equations (5) and (6) yield

$$\delta = -\tfrac{1}{2}\int_0^\tau \omega_L(t)\cos\gamma(t)\,\mathrm{d}t, \qquad \beta = -\tfrac{1}{2}\alpha \tag{7}$$

That is, δ is the accumulated phase change owing to the component of the angular velocity of the Larmor precession in the spin direction, and β is acquired by the additional rotation owing to parallel transport, as explained earlier. More generally, equation (5) gives the accumulated phase change owing to the instantaneous change in $|\psi(t)\rangle$ in the direction of $|\psi(t)\rangle$, according to Schrödinger's equation, while β is the additional phase due to parallel transport. But the parallel-transported state vector undergoes the holonomy transformation, giving $|\psi\rangle$ an additional phase according to equation (6). It has been shown that[2,3], in a sense, the Aharonov–Bohm effect is a manifestation of this phase.

Adiabatic approximation. An important special case of the geometric phase results from the adiabatic cyclic evolution studied by Berry[2]. Suppose that a hamiltonian H is a function of a set of parameters $\mathbf{R}$. By varying $\mathbf{R}(t)$ sufficiently slowly with time t, $H(\mathbf{R})$ can be made to vary adiabatically. The adiabatic theorem in quantum mechanics states that if a quantum state is initially an eigenstate of $H(\mathbf{R}(0))$, it continues to remain as the eigenstate of the 'snapshot' hamiltonian $H(\mathbf{R}(t))$:

$$H(\mathbf{R}(t))|\psi(t)\rangle = E(t)|\psi(t)\rangle \tag{8}$$

In particular, if $\mathbf{R}$ moves around a closed curve so that $\mathbf{R}(0) = \mathbf{R}(\tau)$, the initial and final states are eigenstates of $H(\mathbf{R}(0))$. If $H(\mathbf{R}(t))$ is non-degenerate (that is, the eigenstate is unique up to a phase factor), then $|\psi(\tau)\rangle = e^{i\phi}|\psi(0)\rangle$. The evolution is then cyclic, with $\phi = \beta + \delta$, and from equations (5) and (8), $\delta = -\hbar^{-1}\int_0^\tau E(t')\,\mathrm{d}t'$ in this special case.

Even though adiabaticity is not essential for a geometric phase to be acquired, there are important examples of adiabatic evolution in physics. It occurs, for example, in molecular physics, which forms the basis of all of chemistry. The electronic states of a vibrating or rotating molecule can be determined by calculating the states for stationary nuclei and allowing these to evolve adiabatically as the nuclei move; the coordinates of the electrons and nuclei are called the fast and slow variables respectively. This is the Born–Oppenheimer approximation. As the nuclei travel around a closed path in the parameter space of slow variables, the electronic wavefunction acquires a geometric phase[16–19].

Geometric angles

I will now generalize the treatment of the geometric phase for the neutron to a particle with arbitrary angular momentum, such as an atom or a molecule. A particle with an angular momentum characterized by the quantum number J (an integer or half-integer) has $2J+1$ independent states, which undergo a change of phase when the particle is rotated about some chosen axis, called the quantization axis. These states are denoted by $|n\rangle$, $n = J, J-1, \ldots, -J+1, -J$, where the rotation of the state $|n\rangle$ by an angle θ about this axis results in a phase change of $-n\theta$. For the neutron $J=\frac{1}{2}$ and the two states $n=\frac{1}{2}, -\frac{1}{2}$ correspond to the spin vectors $\mathbf{S}$ and $-\mathbf{S}$ parallel and antiparallel to the quantization axis.

Suppose that the particle is in a magnetic field $\mathbf{B}(t)$ which varies arbitrarily in some interval of time $t=0$ to τ. The effect of this on the spin states is to give all of them the same rotation at every instant of time. Therefore the quantization axis also undergoes a rotation. The net effect induced by the magnetic field during the time interval $t=0$ to τ is a rotation about a unique axis, which must therefore come back to itself. Hence each of the spin states $|n\rangle$ with this axis chosen as the quantization axis undergoes a cyclic evolution, changing only by a phase factor ϕ_n. The possible directions of the quantization axis may be represented by points on a sphere. So the effect of $\mathbf{B}(t)$ in this time interval is to move the quantization axis along a closed curve C on this sphere.

The instantaneous precession of each state $|n\rangle$ about the instantaneous quantization axis is the component of the Larmor

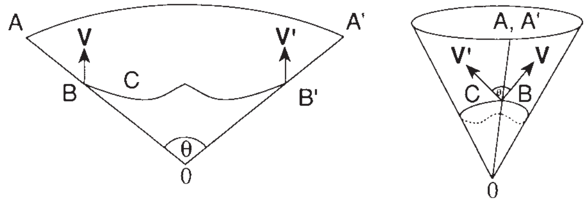

FIG. 4 Parallel transport on a cone as a classical analogy of the Aharonov–Bohm effect. The cone is formed by aligning OA and OA'. The vector **V** parallel-transported from B around the curve C on the cone would come back to B (identified with B') as **V'**, rotated by the angle θ. This is analogous to the Aharonov–Bohm phase shift between beams interfering around C which encloses a magnetic field; the field corresponds to the curvature at the apex of the cone.

frequency in this direction. These accumulate to yield the dynamical phase δ_n, given by equation (7) with n replacing the factor of $\frac{1}{2}$. But as we saw by parallel-transporting a triad around C, each state should come back rotated by the additional angle α, the solid angle subtended by C. Therefore it acquires the additional geometric phase $\beta_n = \phi_n - \delta_n = -n\alpha$. This β_n can also be obtained directly from equation (6).

The angle α is purely geometrical and is independent of the quantum numbers J or n, unlike the geometric phase which is proportional to n. For a more general physical system, however, the evolution of the system cannot be understood as an ordinary rotation. Instead, α generalizes to a set of angles $\alpha_1, \alpha_2, \ldots, \alpha_m$, for quantum systems[6,7] as well as classical systems[7,9]. The classical case can be obtained as the limiting case of the states $|n\rangle$ being so-called WKB (this stands for Wentzel, Kramers and Brillouin) states, which have fairly well defined momenta or actions and therefore poorly defined positions because of the Heisenberg uncertainty relation. These states can be associated with ensembles of classical particles with definite momentum or action. In this limit, the angles α_k are the same as Hannay's angles[9,10] and their non-adiabatic generalization[7,11], which determine the geometric part of the cyclic evolution undergone by a classical system. Interesting physical applications of these angles in classical physics have been described by Hannay[9] and Berry and Hannay[11].

Geometric phase from measurements

According to the 'Copenhagen interpretation' of quantum mechanics, a quantum state undergoes two types of change. Between measurements it undergoes a continuous evolution governed by the Schrödinger equation; but when a measurement is made, it undergoes a discontinuous change into an eigenstate of the observable being measured. Pancharatnam's pioneering work[5] amounted to identifying a geometric phase in the evolution of a system of photons, caused by the second type of change.

Consider a beam of polarized light, which consists of photons in a given spin state, passing through a sequence of polarizers such that the final polarizer restores the initial polarization. The photons then undergo a cyclic evolution (Fig. 5). Each polarizer acts as a 'measuring device' and the polarization of a photon that passes through it undergoes a discontinuous change. For a given momentum, a photon may be regarded as having two possible independent polarization states, which define a two-dimensional Hilbert space analogous to the spin state of the neutron (although the photon has spin 1). Therefore, the projective Hilbert space is again a sphere, called the Poincaré sphere, points on the surface of which represent the polarization states (without defining their phase). The successive polarization states of the photon correspond to a sequence of points on this sphere.

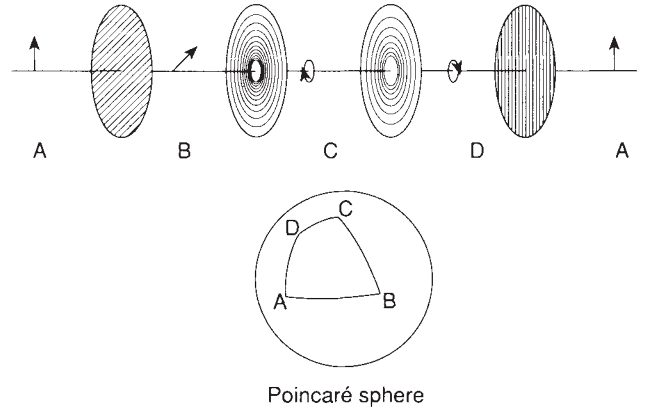

FIG. 5 Pancharatnam's optical experiment which first introduced the geometric phase[5]. Polarized light passes through a sequence of polarizers each of which lets through only the component of light with a particular polarization. A photon therefore goes through a sequence of polarization states represented by the points A, B, C, . . . , A on the Poincaré sphere. The final polarizer restores the initial polarization; but a net phase is acquired by the photon equal in magnitude to half the solid angle of the geodesic polygon ABC, . . . , A on the sphere.

When a photon passes through each polarizer, the photon's state vectors immediately before and after the passage have the same phase as defined earlier, which was shown by Pancharatnam[5]. Berry[12] showed that this is equivalent to parallel-transporting, with respect to the connection defined by equation (3), the photon polarization state vector along the shortest geodesic joining the two points on the Poincaré sphere corresponding to the initial and final states. Therefore for a cyclic evolution the state vector undergoes the holonomy transformation obtained by parallel-transporting it around the geodesic polygon formed from the sequence of points on the Poincaré sphere; and so the photon state acquires a phase of magnitude $\alpha/2$, where α is the solid angle of the geodesic polygon—in agreement with Pancharatnam's result[5]. This result has been generalized to cyclic evolution of an arbitrary quantum system induced by measurements, in which case the geometric phase is obtained by parallel-transporting around the geodesic polygon defined by the sequence of states of the system in the corresponding projective Hilbert space[4,14].

Experimental tests

The acquisition of a geometric phase by a spin has been verified experimentally by Bitter and Dubbers[20] for the special case of adiabatic evolution. They passed a beam of spin-polarized neutrons through a solenoid such that, in the rest frame of the neutron, the magnetic field **B** slowly swept out a cone of solid angle α. The neutron spin states parallel and antiparallel to **B** acquire, in addition to their dynamical phases, the Berry phases $\pm\frac{1}{2}\alpha$. This should result in an additional precession of the neutron by the angle α (the difference between the Berry phases), as Bitter and Dubbers observed. A similar experiment was performed by Suter *et al.*[21], who used techniques of nuclear magnetic resonance (NMR) spectroscopy to subject a spin to an effective magnetic field which slowly sweeps out a cone.

An ingenious experiment was performed by Tycko[22], who used a crystal of sodium chlorate in which the spins of the chlorine nuclei were aligned with the crystal's symmetry axis, which is here the quantization axis. He excited these nuclei into a superposition of two spin states by using a pulse of radio waves. By rotating the crystal about some other axis, the two spins states then acquired geometric phases that increased with time, producing a difference in the frequency (change of phase per unit time) of the two states. This shift was detected by a change in the frequency of the radiation emitted by the nuclei, measured using standard NMR phase-sensitive detection techniques. The use of mechanical rotation in this experiment makes it clear that the geometric phase is associated with the motion of the system and is independent of the particular hamiltonian inducing the motion[3,4].

The non-adiabatic geometric phase was also observed by Suter *et al.*[23] using NMR techniques. They studied a system of coupled protons, a quantum system with total spin $J = 1$. As seen previously, there are three independent states so that the Hilbert space $\mathscr{H}$ is three-dimensional. A two-level subspace was made to undergo a cyclic evolution in $\mathscr{H}$ by applying a time-dependent magnetic field. The geometric phase was measured by looking at interference between these two states and the third, unperturbed level, and the prediction of Aharonov and Anandan[3] was confirmed for different circuits in $\mathscr{H}$.

An interesting optical experiment was performed by Chiao and colleagues[24,25], who sent polarized light through a helically twisted optical fibre so that the direction of light on entry to and exit from the fibre was the same. As the light passes through the fibre its direction traces out a closed curve C over the sphere of all possible directions. On exit, the plane of polarization is rotated relative to its initial state by the solid angle α subtended by C at the centre of the sphere. This is because the photons, for polarized light, are in a superposition of spin states parallel and antiparallel to their momentum which acquire geometric phases $\mp\alpha$. This corresponds to a rotation of the photon spin (polarization) by the angle α, because the photon has spin 1.

This experiment can also be understood as a special case of the general treatment of Anandan and Stodolsky[6], described earlier. Consider a cartesian triad with its z axis in the direction of the momentum (along the axis of the fibre), which is moved along this axis without rotation about it. With respect to this triad, the plane of polarization does not change[3,26-28]. But the triad can be regarded as undergoing parallel transport along C and therefore comes back rotated by α about its z axis, and the plane of polarization is rotated likewise. A quantum-mechanical description is not needed to understand this experiment because the geometry is so general as to apply to both classical and quantum systems. Again the geometric phase arises entirely from the motion of the system, with adiabaticity being irrelevant.

For further discussion of the diverse ramifications and experimental applications of geometric phases see refs 29 and 30.

Conclusion

The purpose of physics is to simplify and unify our understanding of nature. Geometric phases and angles provide a unified description of a wide range of phenomena in classical and quantum physics. But their greatest value lies perhaps in giving us a new way of looking at quantum theory. It is useful to look at a theory from different viewpoints, because it may be easier to see how to modify the theory from one perspective than from another. Historically, this has been particularly true of geometric reformulations of physical theories. It may not be unreasonable to hope, therefore, that the new insights gained in the past few years through the study of the geometric phase and related geometric structures in quantum mechanics may have heuristic value. □

Jeeva Anandan is in the Department of Physics and Astronomy, University of South Carolina, Columbia, South Carolina 29208, USA. This work was performed partially at the Department of Theoretical Physics, University of Oxford, Oxford OX1 3NP, UK.

1. Aharonov, Y. & Bohm, D. *Phys. Rev.* **115,** 485-491 (1959).
2. Berry, M. V. *Proc. R. Soc.* A**392,** 45-57 (1984).
3. Aharonov, Y. & Anandan, J. *Phys. Rev. Lett.* **58,** 1593-1596 (1987).
4. Anandan, J. & Aharonov, Y. *Phys. Rev.* D**38,** 1863-1870 (1988).
5. Pancharatnam, S. *Proc. Indian Acad. Sci.* A**44,** 247-262 (1956); reprinted in *Collected Works of S. Pancharatnam* (Oxford Univ. Press, 1975).
6. Anandan, J. & Stodolsky, L. *Phys. Rev.* D**35,** 2597-2600 (1987).
7. Anandan, J. *Phys. Lett.* A**129,** 201-207 (1988).
8. Kugler, M. & Shtrikman, S. *Phys. Rev.* D**37,** 934-937 (1988).
9. Hannay, J. H. *J. Phys.* A**18,** 221-230 (1985).
10. Berry, M. V. *J. Phys.* A**18,** 15-27 (1985).
11. Berry, M. V. & Hannay, J. H. *J. Phys.* A**21,** L325-L331 (1988).
12. Berry, M. V. *J. mod. Opt.* **34,** 1401-1407 (1987).
13. Ramaseshan, S. & Nityananda, R. *Current Science* **55,** 1225-1226 (1986).
14. Samuel, J. & Bhandari, R. *Phys. Rev. Lett.* **60,** 2339-2342 (1988).
15. Simon, B. *Phys. Rev. Lett.* **51,** 2167-2170 (1983).
16. Herzberg, G. & Longuet-Higgins, H. C. *Disc. Farad. Soc.* **35,** 77-82 (1963).
17. Stone, A. J. *Proc. R. Soc. Lond.* A**351,** 141-150 (1976).
18. Mead, C. A. & Truhlar, D. G. *J. chem. Phys.* **70,** 2284-2296 (1979).
19. Moody, J., Shapere, A. & Wilczek, F. in *Geometric Phases in Physics* (eds Shapere, A. & Wilczek, F.) 160-183 (World Scientific, Singapore, 1989).
20. Bitter, T. & Dubbers, D. *Phys. Rev. Lett.* **59,** 251-254 (1987).
21. Suter, D., Chingas, G. R. Harris & Pines, A. *Molec. Phys.* **61,** 1327-1340 (1987).
22. Tycko, R. *Phys. Rev. Lett.* **58,** 2281-2284 (1987).
23. Suter, D., Mueller, K. T. & Pines, A. *Phys. Rev. Lett.* **60,** 1218-1220 (1988).
24. Chiao, R. Y. & Wu, Y. *Phys. Rev. Lett.* **57,** 933-936 (1986).
25. Tomita, A. & Chiao, R. Y. *Phys. Rev. Lett.* **57,** 937-940 (1986).
26. Berry, M. V. *Nature* **326,** 277-278 (1987).
27. Rytov, S. M. *Dokl. Akad. Nauk.* **XVIII,** 263 (1938); reproduced in *Topological Phases in Quantum Theory* (eds Markovski, B. & Vinitsky, S. I.) (World Scientific, Singapore, 1989).
28. Haldane, F. D. M. *Phys. Rev. Lett.* **59,** 1788 (1987).
29. *Geometric Phases in Physics* (eds Shapere, A. & Wilczek, F.) (World Scientific, Singapore, 1989).
30. Zwanziger, J. W., Koenig, M. & Pines, A. *A. Rev. phys. Chem.* **41,** 601-646 (1990).

α-Inhibin is a tumour-suppressor gene with gonadal specificity in mice

Martin M. Matzuk*†, Milton J. Finegold†, Jyan-Gwo J. Su‡, Aaron J. W. Hsueh‡ & Allan Bradley*

* Institute for Molecular Genetics and † Department of Pathology, Baylor College of Medicine, Houston, Texas 77030, USA
‡ Division of Reproductive Biology, Department of Gynecology/Obstetrics, Stanford University, Stanford, California 94305, USA

The inhibins are α:β heterodimeric growth factors that are members of the transforming growth factor-β family. To understand the physiological roles of the inhibins in mammalian development and reproduction, a targeted deletion of the α-inhibin gene was generated by homologous recombination in mouse embryonic stem cells. Mice homozygous for the null allele (inhibin-deficient) initially develop normally but every mouse ultimately develops mixed or incompletely differentiated gonadal stromal tumours either unilaterally or bilaterally. Inhibin is thus a critical negative regulator of gonadal stromal cell proliferation and the first secreted protein identified to have tumour-suppressor activity.

THE inhibins and activins are developmentally and physiologically important dimeric growth factors[1-4] with structural homology to a large group of proteins including the transforming growth factor (TGF)-βs (ref. 5), Mullerian-inhibiting substance (MIS)[6], and multiple proteins related to the *Xenopus* Vg1 protein and the *Drosophilia* DPP-C protein[7]. Similar to other members of the TGF-β family, these proteins are secreted from cells as dimers and interact with cell surface receptors[8]. The inhibins are $\alpha:\beta$ heterodimers, whereas the activins are $\beta:\beta$ dimers. There are two unique but homologous β-subunits, βA and βB, which are shared between these growth factors (Fig. 1a)[1].

The inhibins and activins were initially discovered as gonadal peptides that either inhibited or stimulated pituitary follicle stimulating homone (FSH) production, respectively[1]. Several studies suggest, however, that the inhibins and activins may have critical intragonadal paracrine and/or autocrine roles, often acting as mutual antagonists. The major gonadal sites of inhibin synthesis are the Sertoli cells in males and the granulosa