Lecture notes for
The Quantum Hall Effect and Beyond:
The Integer Quantum Hall Effect

W. Zhu[1]

[1] *Westlake Institute of Advanced Study,*

*Westlake University, Hangzhou, P. R. China**

Contents

[4]Reporting of typos, inaccuracies and errors to zhuwei@westlake.edu.cn would be greatly appreciated.

INTRODUCTION

This course will address the topics related to the quantum Hall effect. Personally I think it is a very important subject, which should be an introductary level course to graduate students in Condensed Matter Physics. The reason is many-fold. First of all, the quantum Hall physics is one of two most important discoveries in physics, since 1980s. The concepts, ideas and knowledges reflected in this area totally reshaped the modern physics especially condensed matter physics. It changes the way that people think about problems. Moreover, the discovery of quantum Hall effect gives birth to the topological physics. Nowadays, topology is one of most exciting field in condensed matter physics. Without any doubt, the physics related to topology is one of most exciting field in condensed matter physics since 1980s. There are in total three Nobel prizes related to this topic: the discovery of integer quantum Hall effect (1980, Nobel prize in 1985), the discovery of fractional quantum Hall effect (1982, Nobel prize in 1998), the topolgocal states of matter (1974, 1982, 1984, 1988, Nobel prize in 2016). We will try to cover these progresses in this lecture.

The basic knowledges in the quantum Hall effect are important to understand where the modern condensed matter physics comes from, why condensed matter physics becomes "condensed matter", and how it looks like now. So these knowledges should be the ground for a graduate student of condensed matter physics.

Topological states of matter is a surprising outcome of condensed matter physics, where systems that are full of details give rise to physical phenomena that are astonishingly independent of many irrelevant setup details. The most striking example, the quantum Hall effect, shows a quantization of the Hall resistivity to a level within error around 10^{-9}, with the quantum being as universal as physics gets, h/e^2. Soonafterwards, the fractional quantum Hall effect exposed the way in which interaction between electrons may lead to wide spread fractionalization - of the charge, the spin, the statistics, the central charge, and of several response functions. In 2005, the discovery of the time-reversal-symmetric analogs of the quantum Hall effect, the topological insulators, comes into the condensed matter physics.

Nowadays, topolgocal states of matter has been spaned to a wide family. Here we just mention a general framework. A naive definition of topological states of matter: Some of physical properties are robust due to intrinsic reasons (e.g. symmetries), and usually these properties can be described by some topological quantum numbers, which are independent

of details of setup. Many of interested states can be grouped into the following table or graph.

TABLE I: A general classfication of topological states of matter

Topological States of Matter	
Topological Band Structure	Intrinsic Topological Order
Integer quantum Hall Effect	Fractional Quantum Hall Effect
Quantum Anomalous Hall Effect (Chern Insulator)	Fractional Chern Insulator
Quantum Spin Hall Effect	Quantum Spin Liquid
Topological Insulator	Fractional Topological Insulator
Topological Crystaline Insulator	
Topological Semimetal	

The quantum Hall physics, in the center stage of topological states of matter, has become an extremely important research subject during the last two and a half decades. The interest for quantum Hall physics stems from the low- dimensional quantum systems to strongly-correlated systems, and probably covers many key areas of modern condensed-matter physics. From a theoretical point of view, the study of quantum Hall systems brings many novel concepts some of which were better known in quantum-field theories used in high-energy, such e.g. charge fractionalisation, non-commutative geometries and topological field theories. And one will see many exotic techniques and tools for topological states that is beyond the standard solid state physics. Thus, I choose it as the main focus of this lecture.

The motivation of the present lecture notes is to provide in an accessible manner the basic knowledge of the quantum Hall physics and topological states of matter. I hope it will let interested graduate students to pursue on their own further studies in this subject. I feel that a more detailed discussion of some aspects in this large field of physics would go beyond the introductory character of this lecture note. Finally, through this lecture, I hope to clarify and answer the following questions:

- What do we mean by "Insulators"?

- What do we mean by "Topology" ?
- What do we mean by "Topological Insulators" ?
- What do we mean by "Fractional" ?

I hope these questions will be clear after this lecture is done.

PRELIMINARY

Electric and mangetic field

These are two most common field in physics, which are widely studied in the past centuries. An electric field is generated by a voltage, and a magnetic field is generated by magnet. Their relationship can be clearly explained with the help of Maxwell's Equations, a set of partial differential equations which relate the electric and magnetic fields to their sources, current density and charge density.

$$Gauss's law : \oint_S \mathbf{E} \cdot ds = Q \tag{1}$$

$$Gauss's law for magnet : \oint_S \mathbf{B} \cdot ds = 0 \tag{2}$$

$$Faraday's law : \oint_C \mathbf{E} \cdot dl = -\frac{d}{dt} \int_S \mathbf{B} \cdot ds \tag{3}$$

$$Ampre - Maxwell's law : \oint_C \mathbf{B} \cdot dl = \mu_0 [I_0 + \frac{d}{dt} \int_S \mathbf{E} \cdot ds] \tag{4}$$

Gauss's law: The electric flux through any closed surface is equal to the electric charge Q enclosed by the surface. Gauss's law describes the relation between an electric charge and the electric field it produces. This is often pictured in terms of electric field lines originating from positive charges and terminating on negative charges, and indicating the direction of the electric field at each point in space.

Gausss law for magnetism: The magnetic field flux through any closed surface is zero. This is equivalent to the statement that magnetic field lines are continuous, having no beginning or end. Any magnetic field line entering the region enclosed by the surface must also leave it. No magnetic monopoles, where magnetic field lines would terminate, are known to exist

Faradays law: A changing magnetic field induces an electromotive force and, hence, an electric field. The direction of the force opposes the change. Changing magnetic flux through a surface induces an electromotive force (EMF) in any boundary path of that surface. A changing magnetic field induces a circulating electric field. The voltage accumulated around a closed circuit is proportional to the time rate of change of the magnetic flux it encloses.

Ampre's law: An electric current I or a changing electric flux through a surface produces a circulating magnetic field around any path that bounds that surface. Electric currents

and changes in electric fields are proportional to the magnetic fields circulating about the areas where they accumulate.

We can also express the Maxwell equations in the differation form:

$$\partial \cdot \mathbf{E} = \rho \tag{5}$$

$$\partial \cdot \mathbf{B} = 0 \tag{6}$$

$$\partial \times \mathbf{E} = -\frac{d}{dt}\mathbf{B} \tag{7}$$

$$\partial \times \mathbf{B} = \mu_0[\mathbf{J} + \frac{d}{dt}\mathbf{E}] \tag{8}$$

The usual gauge choice is like following. $\mathbf{E} = -\partial\phi - \frac{\partial \mathbf{A}}{\partial t}$ and $\mathbf{B} = \partial \times \mathbf{A}$. This choice satisfies the Maxwell equation.

When an electron moving in a electric and magnetic field, the Lorentz force is

$$\mathbf{F} = q(\mathbf{E} + \mathbf{v} \times \mathbf{B}). \tag{9}$$

In an electromagnetic field, the directions in which the electric and magnetic field move, are perpendicular to each other.

The classical Hall effect

The behaviour of charged particles moving in a magnetic field has been of interest to physicists for many decades. In 1879, E. H. Hall carried out the first measurements of the current flowing through a thin metallic strip in the presence of a perpendicular magnetic field, using the set-up shown in Fig. 1. Remarkably, he detected a current that flows in a perpendicular direction to the applied voltage, and whose strength is proportional to the applied magnetic field and the longitudinal current.

This behaviour was attributed to the Lorentz force, $F = q[E + v \times B]$, which acts on a particle with charge q travelling with velocity v in the presence of an electric field, E, and magnetic field, B. In free (two-dimensional) space, the presence of a perpendicular magnetic field causes moving charged particles to form circular cyclotron orbits, oscillating at the cyclotron frequency, $\omega_c = |q|B/m$. [Classically, because of the electrons move always perpendicular to the magnetic field the Lorentz force can be simplified to $F = ev_0B$. You can get the radius because the Lorentz force is here the centripetal force $F = \frac{mv_0^2}{r}$, with the

electron mass m and the radius r. Then, $r = \frac{mv_0}{eB}$. So the period is $T = \frac{2\pi r}{v_0} = 2\pi\frac{m}{eB}$.] In a classical two-dimensional slab of metal, however, deflected electrons near the edge cannot complete circular orbits, and so charge starts to build up at the side of the sample. In turn, this creates a transverse voltage and electric field, which eventually opposes the Lorentz force. When these forces are balanced, the equilibrium potential difference is known as the Hall voltage (V_H), and leads to a current in a four-terminal measurement.

The classical Hall effect refers to the phenomenon whereby a current carrying conductor (with current I) exposed to a external magnetic field B, develops a transverse potential difference V_H (and conversely). We define the Hall resistance as

$$R_H = \frac{V_H}{I} = \frac{b \cdot E}{b \cdot j} = \frac{vB}{qn_e v} = \frac{B}{qn_e} \tag{10}$$

where q is the charge of the current carriers and n_e is the number of carriers per unit area.

Remark. 1. R_H is large when n_e is small, so that thin conductors generate large Hall resistance.

2. By measuring R_H we can determine the sign of q and thus determine whether current is carried by electrons or holes.

More quantitatively, the classical Hall effect may be understood within the Drude model for diffusive transport in a metal. In this model, we consider independent charge carriers of momentum $\mathbf{p}$ described by the equation of motion

$$\frac{d\mathbf{p}}{dt} = -e(\mathbf{E} + \frac{\mathbf{p}}{m} \times \mathbf{B}) - \frac{\mathbf{p}}{\tau} \tag{11}$$

The last term takes into account relaxation processes due to the diffusion of electrons by generic impurities, with a characteristic relaxation time τ. The macroscopic transport characteristics, i.e. the resistivity or conductivity of the system, are obtained from the static solution of the equation of motion, $dp/dt = 0$, and one finds

$$eE_x = -\frac{eB}{m}p_y - p_x/\tau \tag{12}$$

$$eE_y = \frac{eB}{m}p_x - p_y/\tau \tag{13}$$

where we have assumed the magnetic field is along z-direction. The cyclotron grequency is $\omega_c = \frac{eB}{m}$. With the help of the Drude conductivity, $\sigma_0 = \frac{ne^2\tau}{m}$, we have

$$\sigma_0 E_x = -enp_x/m - enp_y/m(\omega_c\tau) \tag{14}$$

$$\sigma_0 E_y = enp_x/m(\omega_c\tau) - enp_y/m \tag{15}$$

or in terms of the current density $\mathbf{j} = -en\mathbf{p}/m$, and in the matrix form as $\mathbf{E} = \rho\mathbf{j}$, the resistivity tensor is

$$\rho = \sigma^{-1} = \frac{1}{\sigma_0}\begin{bmatrix} 1 & \omega_c\tau \\ -\omega_c\tau & 1 \end{bmatrix} = \frac{1}{\sigma_0}\begin{bmatrix} 1 & -\mu B \\ -\mu B & 1 \end{bmatrix} \tag{16}$$

and the mobility is $\mu = \frac{e\tau}{m}$. From the above expression, one may immediately read off the Hall resistivity $\rho_H = \frac{\omega_c\tau}{\sigma_0} = \frac{B}{en}$. Furthermore, we have conductivity tensor is

$$\sigma = \rho^{-1} = \sigma_0\begin{bmatrix} 1/(1+\omega_c^2\tau^2) & -\omega_c\tau/(1+\omega_c^2\tau^2) \\ \omega_c\tau/(1+\omega_c^2\tau^2) & 1/(1+\omega_c^2\tau^2) \end{bmatrix} \tag{17}$$

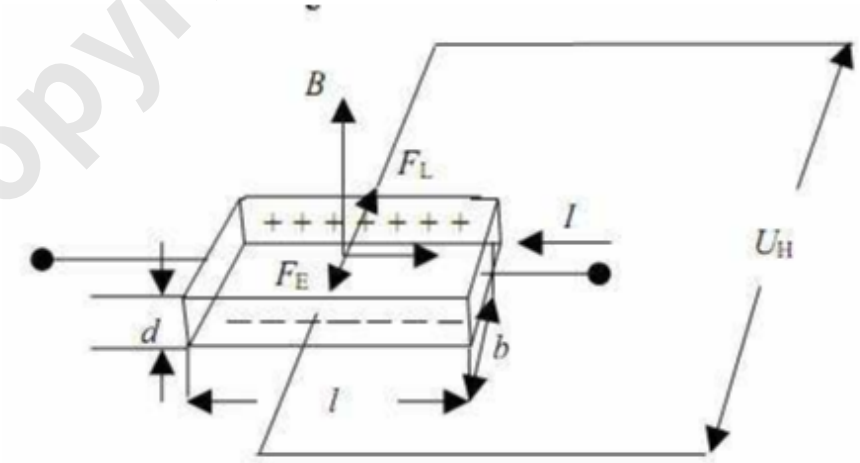

FIG. 1: The classical Hall Effect.

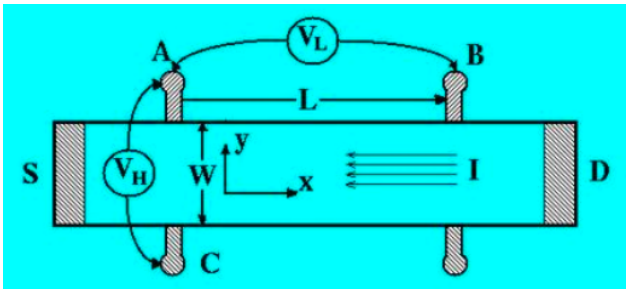

FIG. 2: The Hall bar setup.

EXPERIMENTAL OBSERVATIONS

EXPERIMENTAL SETUP

Hall bar

The Hall bar setup is shown in Fig. 2. The experimental sample is placed in the center. A current I is sent through the material, along the direction labeled x. The voltage change Vx is measured between the two leads on top. This configuration is a standard four-probe measurement. The voltage contacts are not put at the end of the sample, in order to avoid the influence of nonohmic contacts where the current enters and leaves. Another pair of voltage contacts measure the voltage V_y across the sample. This voltage is zero in the absence of a magnetic field.

Resistance and resistivity

We have the relation:

$$\begin{bmatrix} E_1 \\ E_2 \end{bmatrix} = \begin{bmatrix} \rho_{11} & \rho_{21} \\ \rho_{12} & \rho_{22} \end{bmatrix} \begin{bmatrix} j_1 \\ j_2 \end{bmatrix} = \begin{bmatrix} \rho_{11} & \rho_{21} \\ \rho_{12} & \rho_{22} \end{bmatrix} \begin{bmatrix} I/W \\ 0 \end{bmatrix} = \frac{I}{W} \begin{bmatrix} \rho_{11} \\ \rho_{21} \end{bmatrix} \tag{18}$$

So we have the longitudinal voltage is

$$V_L = E_1 L = \rho_{11} \frac{L}{d} I, \quad R_L = \rho_{11} \frac{L}{d} \tag{19}$$

and the transverse voltage is

$$V_H = E_2 W = \rho_{21} \frac{I}{W} W = \rho_{21} I, \quad R_H = \rho_{21} \tag{20}$$

In two dimensional Hall bar, we have that the resistance R_H is equal to the resistivity ρ. This relation is important for experiments.

Two-dimensional electron gas

As already mentioned above, the history of the quantum Hall effect is related to technological advances in the fabrication of 2D electron systems with high electronic mobilities. The increasing mobility allows one to probe the fine structure of the Hall curve and thus to observe those quantum Hall states which are more fragile, such as some exotic FQHE states. In this sense, electronic mobility means resolution and the tiny object is the quantum Hall state. As an order of magnitude, todays best 2D electron gases (in GaAs/AlGaAs heterostructures) are characterised by mobilities $\mu \sim 10^7 cm^2/Vs$.

Remark. 1) high mobility; 2) 2D nature.

MOSFET

The samples used in the discovery and in the first studies of the IQHE were metal-oxide-semiconductor field-effect transistors (MOSFET). A metallic layer is seperated from a semiconductor (typically doped silicon) by an insulating oxide (e.g. SiO2) layer (see inset I in Fig.). The chemical potential in the metallic layer may be varied with the help of a gate voltage V_G. At $V_G = 0$, the Fermi energy in the semiconductor lies in the band gap below the acceptor levels of the dopants [Fig. 3(a)]. When lowering the chemical potential in the metal with the help of a positive gate voltage $V_G > 0$, one introduces holes in the metal that attract, via the electric field effect, electrons from the semiconductor to the semiconductor-insulator interface. These electrons populate the acceptor levels, and as a consequence, the semiconductor bands are bent downwards when they approach the interface, such that the filled acceptor levels lie now below the Fermi energy (Fig. 3(b)). Above a certain threshold of the gate voltage, the bending of the semiconductor bands becomes so strong that not only the acceptor levels are below the Fermi energy, but also the conduction band in the vicinity of the interface which consequently gets filled with electrons [Fig.3 (c)]. One thus obtains the electrons in the conduction band, the dynamics of which is quantised into discrete electronic subbands in the perpendicular z-direction (see inset II in Fig. 3). Naturally, the electronic

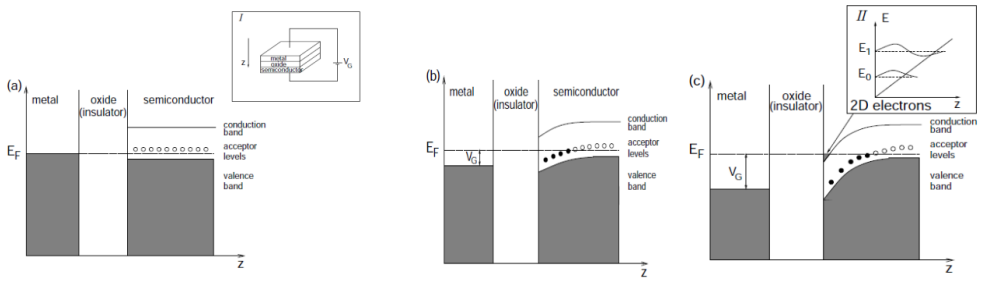

FIG. 3: The inset I shows a sketch of a MOSFET. (a) Level structure at VG = 0. In the metallic part, the band is filled up to the Fermi energy EF whereas the oxide is insulating. In the semiconductor, the Fermi energy lies in the band gap (energy gap between the valence and the conduction bands). Close to the valence band, albeit above EF , are the acceptor levels. (b) The chemical potential in the metallic part may be controled by the gate voltage VG via the electric field effect. As a consequence of the introduction of holes the semiconductor bands are bent downwards, and above a threshold voltage (c), the conduction band is filled in the vicinity of the interface with the insulator. One thus obtains a 2D electron gas. Its levels (electronic subbands) are represented in the inset II.

wave functions are then extended in the z-direction, but in typical MOSFETs only the lowest electronic subband is filled, such that the electrons are purely 2D from a dynamical point of view, i.e. there is no electronic motion in the z-direction.

Semiconductor heterostructures

The mobility in MOSFETs, which is typically on the order of $\mu \sim 10^6 cm^2/Vs$, is limited by the quality of the oxide-semiconductor interface (surface roughness). This technical difficulty is circumvented in semiconductor heterostructures most popular are GaAs/AlGaAs heterostructures which are grown by molecular-beam epitaxy (MBE), where high-quality interfaces with almost atomic precision may be achieved, with mobilities on the order of $\mu \sim 10^7 cm^2/Vs$.

We put a junction of AlGaAs (with ratio 4 : 1 for the Aluminium) and GaAs together. The electrons are bound to the interface so we obtain a two-dimensional electron gas (2DEG). The gas is bound on the interface but free along it. To understand the interface better, we first picture the two materials separately. The materials are set up in such a way that the

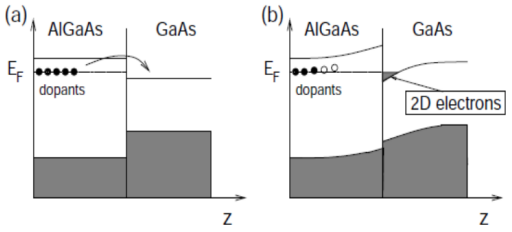

FIG. 4: Semiconductor heterostructure (GaAs/AlGaAs). (a) Dopants are introduced in the AlGaAs layer at a certain distance from the interface. The Fermi energy lies below the in the band gap and is pinned by the dopant levels. The GaAs conduction band has an energy that is lower than that of the dopant levels, such that it is energetically favourable for the electrons in the dopant layer to populate the GaAs conduction band in the vicinity of the interface. (b) This polarisation bends the bands in the vicinity of the interface between the two semiconductors, and thus a 2D electron gas is formed there on the GaAs side.

Fermi energy of AlGaAs is higher than that of GaAs. At zero temperature the Fermi level on the $Al_xGa_{1-x}As$ side lies just above the bound state of the donors, which lies higher than the bottom of the GaAs conduction band. (The Fermi energy is pinned by these donor levels in AlGaAs, which may have a higher energy than the originally unoccupied conduction band in the GaAs part.) Therefore, electrons bound to the donors move into the GaAs conduction band. (it becomes energetically favourable for the electrons in the donor levels to occupy the GaAs conduction band in the vicinity of the interface. As a consequence, the energy bands of AlGaAs are bent upwards, whereas those of GaAs are bent downwards.) This replacement causes electric polarization: the GaAs is charged negatively and the $Al_xGa_{1-x}As$ is charged positively. So when we bring the two media into contact, electrons (assumed with positive charge) spill over from donor sites of AlGaAs, which creates a dipole layer and potential difference. In this way, two-dimensional electrons are formed automatically in this system.

The typical 2D electronic densities in these systems are on the order of $n_e \sim 10^{11} cm^{-2}$, which is much lower than in usual metals. The mobilities is on the order of $\mu \sim 10^7 cm^2/(Vs)$. The high mobility is necessary to observe the FQHE, which was indeed first observed in a GaAs/AlGaAs sample.

Remark: Other 2D electron gas for quantum Hall systems: Graphene; Si/SiGe; GaAlAs/AlAs/GaAlAs; ZnO/MgZnO:

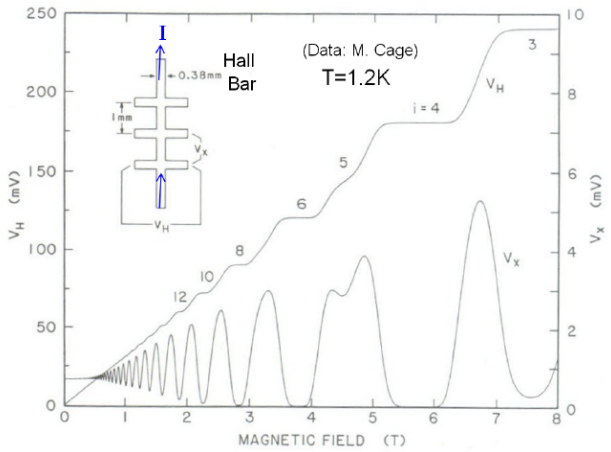

FIG. 5: The experimental observation of IQHE.

Experimental Results

The integer quantum Hall effect (IQHE) is the experimental discovery (von Klitzing, 1980) that the Hall resistance takes integer values.

The experiment exhibits the following features: 1. Plateaus of Hall resistance takes quantization at some interger values.

2. For integer or near integer values, the longitudinal resistance is zero.

3. The width of the plateaus increases with disorder. In fact the disorder will turn out to be crucial for the quantization effect, though too much disorder will destroy it.

4. The accuracy os the quantization is 10^{-8}. Indeed, the value of h/e^2 has been measured and found to be $h/e^2 = 25812.807572\Omega$.

Remark. 1. Low temperature 2. High magnetic field.

THE LANDAU LEVEL PROBLEM

Tips: Landau level. 1929 by Lev Landau

The appearance of the Planck constant h in the formula for R_H is an indication of the inherently quantum mechanical nature of the effect. In this chapter, we study a single electron confined to two dimensions and exposed to a magnetic field. This problem was solved exactly soon after the invention of quantum mechanics (Landau 1929), because it is merely a one-dimensional simple harmonic oscillator problem. The most remarkable aspect of the solution is that the electron kinetic energy is quantized. The discrete kinetic energy levels are called "Landau levels".

Hamiltonian

Our system is free spinless electrons subject to a magnetic field perpendicular to the two dimensional plane. We work in the one-electron approximation, in the one-particle Hilbert space. The Hamiltonian (due to Landau) is given by

$$H = \frac{1}{2m}(\mathbf{p} - e\mathbf{A})^2 \tag{21}$$

where A(r) is the vector potential that generates the magnetic field

$$\mathbf{B} = \nabla \times \mathbf{A} \tag{22}$$

In this equation p is the canonical momentum operator, which satisfies the canonical commutation relation $[p_a, r_b] = -i\hbar\delta_{ab}$. The substitution is also called minimal substitution.

Just as in classical mechanics, the canonical momentum operator is not proportional to the velocity operator. The latter, v, is derived from the Heisenberg equation of motion as follows: $v_a = i[H, r_a] = (\mathbf{p} - e\mathbf{A})_a/m$. We introduce what we call the dynamical momentum operator π_a, which is proportional to the velocity operator: $\pi_a = mv_a = (\mathbf{p} - e\mathbf{A})_a$.

The momentum operator satisfies

$$[\pi_a, \pi_b] = i\hbar eB = -i\frac{\hbar^2}{\ell^2} \tag{23}$$

where the magnetic length $\ell = \sqrt{\hbar/(|e|B)}$. Let us prove it here.

$$\begin{aligned}[\pi_x, \pi_y] &= [\mathbf{p}_x - e\mathbf{A}_x(\mathbf{r}), \mathbf{p}_y - e\mathbf{A}_y(\mathbf{r})] \\ &= e([p_x, A_y] - [p_y, A_x]) = e(\frac{\partial A_x}{\partial x}[p_x, x] + \frac{\partial A_x}{\partial y}[p_x, y] - \frac{\partial A_y}{\partial x}[p_y, x] - \frac{\partial A_y}{\partial y}[p_y, y]) \\ &= i\hbar e(\frac{\partial A_x}{\partial y} - \frac{\partial A_y}{\partial x}) = i\hbar eB\end{aligned}$$

This momentum operator also does not commute with the Hamiltonian. There is big physics behind! This is similar to the Heisenberg relation in spin system, which involves non-commutative relation. This is a fully Quantum result, making it quantum! This non-commutativity between the position and its associated momentum is the origin of the Heisenberg inequality according to which one cannot know precisely both the position of a quantum mechanical particle.

The magnetic length is the fundamental length scale in the presence of a magnetic field. For a estimation, $\ell \approx 26nm/\sqrt{B[T]}$. (T $= Vs/m^2 = J/(Am^2)$, $\hbar \approx 10^{-34}Js, e = 1.6 \times 10^{-19}As$)

Gauge invariance

The choice of the vector potential is not unique for a given magnetic field. This is called a "freedom in the gauge". Now, the Hamiltonian does not depend on the magnetic field directly but on the vector potential. Thus different choices of the vector potential give different results for the wave function. We need to know how the wave function depends on the choice of the vector potential.

For an arbitrary smooth, real function $\xi(r)$, $A'(r)$ obtained from $A'(r) = A(r) + \partial\xi(r)$ and $A(r)$ stand for the same magnetic field (no singularity in $\xi(r)$). Under such gauge transformation, the wave function should be changed to $\psi(r) \to e^{-i\frac{e}{\hbar}\xi(r)}\psi(r)$.

Notice that because A(r) is not gauge invariant, neither is the momentum p. Indeed, the momentum transforms as $p_a \to p_a - e\nabla_a\xi(r)$ under a gauge transformation in order to compensate the transformed vector potential, such that π_a is gauge-invariant.

Eigenvalue problem

For solving the Hamiltonians, it is convenient to use the pair of conjugate operators π_x and π_y to introduce ladder operators in the same manner as in the quantum-mechanical treatment of the one-dimensional harmonic oscillator: $a = \frac{\ell}{\sqrt{2}\hbar}(\pi_x - i\pi_y), a^\dagger = \frac{\ell}{\sqrt{2}\hbar}(\pi_x + i\pi_y)$, which satisfies $[a, a^\dagger] = 1$, the Hamiltonian is rewritten as

$$H = \hbar\omega_B(a^\dagger a + \frac{1}{2}) \tag{24}$$

Thus the energy eigenvalues are discretized into $E_n = \hbar(n + 1/2)$ like a harmonic oscillator. These discrete energy levels are called the "Landau levels". The Landau levels are obtained from the commutation relation of π_a, so they do not depend on the choice of the gauge; this is as it should be.

The number operator $a^\dagger a|n\rangle = n|n\rangle$. The ladder operators act on these states in the usual manner $a^\dagger|n\rangle = \sqrt{n+1}|n+1\rangle$ and $a|n\rangle = \sqrt{n}|n-1\rangle$.

Next we will work out the eigenvalue problem explicitly.

1) We select the Landau gauge

$$\mathbf{A} = B(0, x, 0) \tag{25}$$

the Hamiltonian does not contains y, and therefore commutes with p_y. That implies that $p_y = \hbar k_y$ is a good quantum number:

$$\frac{1}{2m}[-\partial_x^2 + (p_y - eBx)^2]\psi(x, y) = E\psi(x, y) \tag{26}$$

$$\frac{1}{2m}[-\partial_x^2 + (\hbar k_y - eBx)^2]\psi(x) = E\psi(x) \to \frac{1}{2m}[-\partial_x^2 + \frac{1}{\ell^2}(\ell k_y - x/\ell)^2]\psi(x) = E\psi(x) \tag{27}$$

where we set the form of wave function as $\psi(x, y) = e^{ik_y y}\psi(x)$.

In terms of dimensionless quantities

$$x' = x/\ell - \ell k_y \tag{28}$$

$$p'_x = \ell p_x/\hbar \tag{29}$$

we have the hamiltonian as

$$H = \frac{\hbar\omega_B}{2}[(x')^2 + (p'_x)^2], [-\partial_x^2 + x^2]\psi(x) = E/\hbar\omega_B\psi(x) \tag{30}$$

which is the familiar Hamiltonian of a one-dimensional harmonic oscillator. Here, $\hbar\omega_B$ is the cyclotron energy, with $\omega_B = \frac{eB}{m}$.

The eigenvectors are

$$\psi_{n,k_y}(r) = \frac{1}{[\pi 2^{2n}(n!)^2]^{1/4}} \exp[ik_y y] \exp[-\frac{(x-k_y\ell^2)^2}{2\ell^2}] H_n(\frac{(x-k_y\ell^2)}{\ell}) \tag{31}$$

Here we emphasize two points: 1) The energy does not depend on k_y. The eigenstates with different k_y in a given Landau level are degenerate. 2) The x position depends on k_y. An eigenfunction is Gaussian-localized in a narrow strip of width $\sim \ell$ centered at $x_m = k_y\ell^2$. 3) To solve the differential equation for the harmonic oscillator $\frac{\partial^2\psi(x,y)}{\partial^2 x} + (E - x^2)\psi(x,y) = 0$, first notice k_y is a good quantum number, so we use the form for wave function $\psi(x,y) = e^{ik_y y}\psi(x)$. At the second step, we need consider the boundary condition $x \to \infty$ we have $\psi(x) \to 0$, so the form of wave function should be $\psi(x) \sim u(x)e^{-x^2/2}$. The solution of we get $\partial^2 u(x)/\partial^2 x - 2x\partial u(x)/\partial x + (E-1)u(x) = 0$ is $u(x) = H(x)$.

2) The symmetric gauge: $\mathbf{A} = B(-y/2, x/2, 0)$.

$$H = \frac{1}{2}[(-i\partial_x - \frac{y}{2})^2 + (-i\partial_y + \frac{x}{2})^2] = \frac{1}{2}[-4\frac{\partial^2}{\partial z \partial \overline{z}} + \frac{1}{4}z\overline{z} - z\frac{\partial}{\partial z} + \overline{z}\frac{\partial}{\partial \overline{z}}] \tag{32}$$

where we define $z = x - iy = re^{-i\theta}, \overline{z} = x + iy = re^{i\theta}$, $\partial_x = \partial_z + \partial_{\bar{z}}, \partial_y = -i(\partial_z - \partial_{\bar{z}})$. The ladder operators can be defined as

$$a^\dagger = \tfrac{1}{\sqrt{2}}(\tfrac{\bar{z}}{2} - 2\tfrac{\partial}{\partial z}) \tag{33}$$

$$a = \tfrac{1}{\sqrt{2}}(\tfrac{z}{2} + 2\tfrac{\partial}{\partial \bar{z}}) \tag{34}$$

$$b^\dagger = \tfrac{1}{\sqrt{2}}(\tfrac{z}{2} - 2\tfrac{\partial}{\partial \bar{z}}) \tag{35}$$

$$b = \tfrac{1}{\sqrt{2}}(\tfrac{\bar{z}}{2} + 2\tfrac{\partial}{\partial z}) \tag{36}$$

$$[a, a^\dagger] = 1, \quad [b, b^\dagger] = 1 \tag{37}$$

The hamiltonian becomes

$$H = a^+ a + 1/2 \tag{38}$$

In addition, The z component of the angular momentum operator is defined as

$$L_z = -i\hbar\frac{\partial}{\partial\theta} = -\hbar(z\frac{\partial}{\partial z} - \overline{z}\frac{\partial}{\partial \overline{z}}) = -\hbar(b^\dagger b - a^\dagger a) \tag{39}$$

Exploiting the property $[H, L_z] = 0$, the eigenfunctions are chosen to diagonalize H and L simultaneously. The eigenvalue of L is denoted by $m\hbar$; with this definition the quantum number m takes values $-n, -n+1, ...$

The ground state wave function is solved by $a|0,0\rangle = 0, b|0,0\rangle = 0$. We obtain

$$\langle r|0,0\rangle = \frac{1}{\sqrt{2\pi}\ell} e^{-\frac{|z|^2}{4}}. \tag{40}$$

The other wavefunctions are obtained by (The wave function in the lowest Landau level)

$$\langle r|n,m\rangle = \frac{(a^\dagger)^n}{\sqrt{n!}} \frac{(b^\dagger)^m}{\sqrt{m!}} \langle r|0,0\rangle \tag{41}$$

This wave function represents an electron localized circularly as shown in disk. The maximum of the existence probability is on the circumference of a circle of radius $\sqrt{2m}\ell$, and the spread of the wave function in the radial direction is of the order of ℓ. The expectation value of r^2 is $\langle 0,m|r^2|0,m\rangle = 2(m+1)\ell^2$.

Exercises. Prove the normalization condition of electron wave functions.

$$\begin{aligned}
\langle 0,m|0,m'\rangle &= \int \overline{\frac{1}{\sqrt{2\pi 2^m m!}\ell} z^m e^{-|z|^2/4}} \frac{1}{\sqrt{2\pi 2^{m'} m'!}\ell} z^{m'} e^{-|z|^2/4} && (42)\\
&= \frac{1}{\sqrt{2\pi 2^m m!}\ell} \frac{1}{\sqrt{2\pi 2^{m'} m'!}\ell} \int_{\psi=0}^{2\pi} \int_0^\infty dr d\psi e^{-i(m-m')\psi} r^{m+m'+1} e^{-r^2} && (43)\\
&= \frac{\delta_{m,m'}}{m!} m! = \delta_{m,m'} && (44)
\end{aligned}$$

Exercises. Prove the average area of each Landau orbital.

Degeneracy

We have learnt that the energy of 2D charged particles is characterised by a quantum number n, which denotes the Landau level index. But the system hosts large degeneracy for each Landau level.

Semi-classically, the area of each electron orbital is $2\pi\ell^2$. This minimal surface plays the same role as the surface (action) h in phase space and therefore allows us to count the number of possible quantum states of a given (macroscopic) surface S:

$$N_\phi = \frac{S}{2\pi\ell^2}. \tag{45}$$

Notice that the flux density

$$n_B = \frac{1}{2\pi\ell^2} = \frac{B}{e/h} \tag{46}$$

which is nothing other than the magnetic field measured in units of the flux quantum h/e. Therefore, the number of quantum states in a Landau level equals the number of flux quanta.

In order to describe the LL filling it is therefore useful to introduce the dimensionless ratio between the number of electrons N_e and that of the flux quanta

$$\nu = \frac{N_e}{N_\phi} \tag{47}$$

The Landau level degeneracy can be made more precisely. Considering the Landau gauge, we have $k_y = m2\pi/L_y, m \in [1, ..., N_\phi]$ for the wave vector in the x-direction, one may count the number of states in a rectangular surface of length L_y and width L_x. In x-direction, max value of x is also bound by $x_{max} = k_y\ell^2 = N_\phi 2\pi\ell^2/L_y = L_x$, so we have,

$$N_\phi = \frac{L_x L_y}{2\pi\ell^2}. \tag{48}$$

The LL degeneracy can also be obtained readily in the symmetric gauge. Consider a disk of radius R centered at the origin, and ask how many states lie inside it in a given Landau level. Taking, for simplicity, the lowest Landau level, the eigenstate $|0, m\rangle$ has its weight located at the circle of radius $r = \sqrt{2m}\ell$. Thus the largest value of m_{max} for which the state falls inside the disk is given by $m_{max} = R^2/2\ell^2 = \pi R^2/2\pi\ell^2$, which is also the total number of eigenstates in the lowest Landau level that fall inside the disk. Thus, the degeneracy per unit area is $(2\pi\ell^2)^{-1} = eB/h$.

Remark. 1. Gauge. 2. Geometry.

THEORY OF IQHE: LAUGHLIN'S PUMP PICTURE

In order to explain the quantization of the Hall resistivity, Laughlin considered a gedanken experiment as shown in Fig. 6. Here the two-dimensional surface is bent into a cylinder, and electrodes are placed at the edges of the cylinder. The radius of the cylinder is R, and the length is L. A magnetic field of constant magnitude is applied perpendicular to the cylindrical surface. The x, y coordinates are chosen as shown in the figure: the x axis is parallel to the axis of the cylinder, and the y axis is along the circumference. We place a solenoid on the axis of the cylinder, and a magnetic flux Φ is generated inside this solenoid. The solenoid plays an important role in the following argument. (This placement of the solenoid, which is possible in a cold-atom system by "shaking the lattice".) The magnetic field of the solenoid exists only inside the solenoid, and does not exist at the surface of the cylinder. However, the vector potential of this field has a finite value on the surface and affects the motion of electrons. This is the Aharonov-Bohm effect.

The vector potential of the magnetic field penetrating the cylinder surface is given by

$$\mathbf{A} = (0, Bx, 0) \tag{49}$$

and the vector potential generated by the solenoid is

$$\mathbf{A}_\Phi = (0, \Phi/(2\pi R), 0). \tag{50}$$

It can be deduced by

$$\Phi = \int dS \mathrm{rot}\mathbf{A}_\Phi = \int d\mathbf{l} \cdot \mathbf{A}_\Phi = 2\pi R \times \Phi/(2\pi R) \tag{51}$$

(it should since there is no magnetic flux coming from on the cylinder).

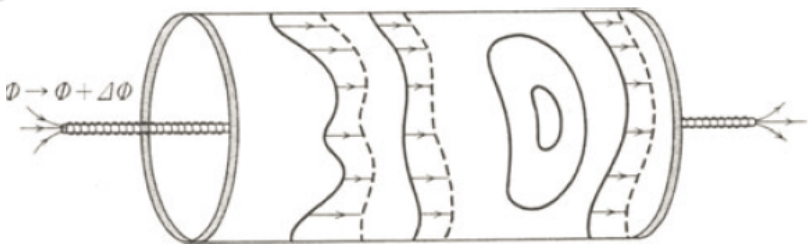

FIG. 6: Gedankenexperiment considered by Laughlin.

Let us see what happens when we change Φ. By Faradays law we know there will be an electric field $\mathbf{E}$ induced along "e" such that

$$\oint_C \mathbf{E} \cdot d\mathbf{l} = -\frac{d\Phi}{dt} \tag{52}$$

where C is a ring around the cylinder in y direction. As a result there will be a Hall current

$$\mathbf{j}_H = \sigma_H \hat{e}_z \times \mathbf{E} \tag{53}$$

Then the current across the fiducial line C is

$$\begin{aligned}
\frac{dQ}{dt} &= I = \int_C \mathbf{j}_H \cdot \hat{e}_x dl && (54)\\
&= \int_C \sigma_H (\hat{e}_z \times \mathbf{E}) \cdot \hat{e}_x dl && (55)\\
&= \sigma_H \int_C (\hat{e}_z \times \hat{e}_x) \cdot \mathbf{E} dl && (56)\\
&= \sigma_H \int_C \hat{e}_y \cdot \mathbf{E} dl && (57)\\
&= -\sigma_H \frac{d\Phi}{dt} && (58)
\end{aligned}$$

Here we get that, the transport rate of charge along the cylinder axis direction is related to the rate of flux, or equivaliently the transport charge is related to the change of flux: $\Delta Q = \sigma_H \Delta\Phi$. Then we consider the flux changes "flux quanta", $\Delta\Phi = h/e$, then we have $\Delta Q/e = \sigma_H(h/e^2)$. Thus we reach the statement of the gedanken experiment: *as* Φ *increases from* Φ *to* $\Phi + 2\pi$ *(a flux quanta), a charge* $\Delta Q = \sigma_H(h/e^2)$ *will be transported (or "pumped") from the left edge to the right edge.*

The Hamiltonian is given by (the difference due to the gauge solenoid and the addition of a flux gauge potential.)

$$\begin{aligned}
H &= \frac{1}{2m}(\mathbf{p} - e\mathbf{A})^2 && (59)\\
&= \frac{1}{2m}(p_x^2 + (p_y - eBx - e\frac{\Phi}{2\pi R}))^2 && (60)\\
&= \frac{1}{2m}(p_x^2 + (\frac{nh}{R} - eBx - e\frac{\Phi}{2\pi R}))^2 && (61)\\
&= \frac{1}{2m}(p_x^2 + (eB)^2(x - \frac{1}{2\pi RB}[n2\pi h/e - \Phi]))^2 && (62)\\
&= \frac{1}{2m}(p_x^2 + (eB)^2(x - x_n(\Phi)))^2 && (63)
\end{aligned}$$

Here we used the quantized condition $p_y = \frac{2\pi\hbar}{R}n$. We recover the harmonic oscillator problem by replacing the cycltron center by $x_n(\Phi)$. So we get the following schematic picture of energy versus the x coordinate: Note that a change of $\Delta\Phi = 2\pi h/e$ for a particular n results in

$$x_n(\Phi+2\pi h/e) = \frac{1}{2\pi RB}[n2\pi h/e-(\Phi+2\pi h/e)]) = \frac{1}{2\pi RB}[(n-1)2\pi h/e-\Phi]) = x_{n-1}(\Phi) \quad (64)$$

so that to change the flux by 2π (flux quanta) is equivalent to merely shifting or relabeling the whole picture n by $n-1$. Thus only what happens at the edges $x = 0$ and $x = L$ will matter, using the fact that occupation is inherited. Thus we see that at the right edge some of the empty states that used to be above the Fermi line will now move below it: charge is "lost" on the right edge. On the left edge, some of the occupied states that used to be below the Fermi line will now move above it: charge is "gained" on the left edge. To compute the total transfer of charge from left to right, we merely have to count the number of occupied Landau levels:

$$\Delta Q = -e \times (\#\text{occupied Landau levels}) \quad (65)$$

which is a quantized number. Thus we have:

$$\sigma_H = \Delta Q\frac{e^2}{h} = (\#\text{occupied Landau levels})\frac{e^2}{h} \quad (66)$$

Note that the number of occupied levels is a deterministic integer, not a random one.

THEORY OF IQHE: LANDAUER PICTURE

Halperin was the first tried to explain the Hall effect by means of the imbalance of the edge currents at opposite edges of the sample. Later, this idea was developed by Buttiker. The Landauer-type and Laughlins derivations of the Hall quantization are closely related. Both require the Fermi level to lie in the localized states in the interior. Both also require extended edge states. Laughlins derivation implicitly assumes that the inner and outer edges are not coupled by disorder, which is equivalent to assuming a suppression of backscattering in the other formulation. The Landauer-type formulation is more general. (i) It more closely resembles experiments. (ii) Laughlins derivation assumes that edge states are phase coherent over the entire length of the sample, which is not the case at nonzero temperatures.

Edge states

The existence of edge states can be motivated classically. In a classical Hall bar with a uniform perpendicular magnetic field, charged particles move in circular motion, the cyclotron orbit. Suppose that our particles are negatively charged and therefore move in counter-clockwise direction. If one places a particle at the right edge of the sample, the particle will do a half-orbit until it collides with the edge. It cant leave the sample and it can also only do counter-clockwise rotational motion, so it will perform another half-orbit, but now it is further down on the edge, see Fig.7. The macroscopic effect is a local current

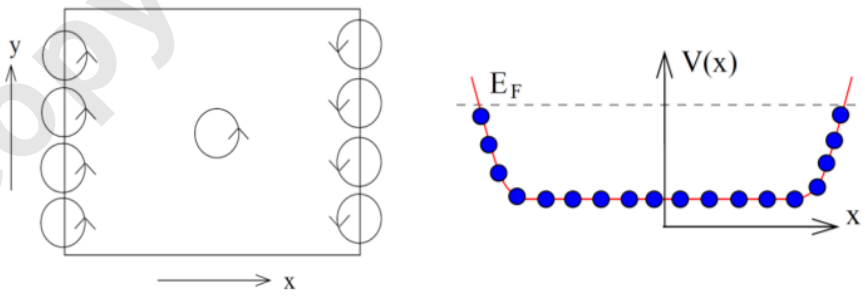

FIG. 7: Classical description of the edge states. Particles at the edge of the sample collide with the edge when moving in the cylcotron orbit and result in an edge current. Model of the samples edges as a steeply rising potential, where the states below the Fermi level are occupied.

on either edge of the sample, with opposing directions, so the net current stays zero.

Edge states can also be treated quantum mechanically. To model the edge of the sample, one introduces a potential well that rises steeply at the edges. Since the states of the unperturbed system are localised in the x-direction, and by assuming that the potential varies slowly on the scale of the magnetic length, we can Taylor expand the potential around these center positions and drop all the constant terms. So the problem we will deal with is like

$$H = \frac{1}{2m}(p_x^2 + (eB)^2(x - k_y\ell^2)^2 + x\frac{\partial V}{\partial x} \tag{67}$$

This Hamiltonian can be solved exactly, by

$$H = \frac{1}{2m}(p_x^2 + (eB)^2(x - k_y\ell^2 + V'(x_n)\frac{1}{(eB)^2})^2 + k_y\ell^2 V'(x_n) - \frac{1}{(eB)^2}V'(x_n)^2 \tag{68}$$

The extra terms are all constant on the k_y momentum subspace so they transfer over to the energy eigenvalues, while the states will have the same form as the ones found earlier. The resulting energy eigenvalues are

$$E_n(k_y) = \hbar\omega_B(n + \frac{1}{2}) - \frac{1}{(eB)^2}V'(x_n)^2 + k_y\ell^2 V'(x_n) \tag{69}$$

They do depend on the y-momentum k_y, so we can calculate a finite group velocity of the wavepackets:

$$v_y = \frac{\partial E}{\hbar\partial k_y} = \frac{1}{eB}V'(x_n) \tag{70}$$

By looking again at the potential picture and using the fact that the states are still localised in x-direction, one sees that the states on the right edge propagate in negative y-direction and opposite for the left side. These edge states are chiral, they can only move in one direction. Since they are chiral, they are immune to back-scattering by impurities. In the classical picture, defects and impurities scatter incoming electrons into random directions, effectively decreasing the current.

Four-terminal

First we consider the ideal two-dimensional electron system shown in Fig. 8. The upper and lower edges are connected to electrodes, and the right and left sides of the sample are

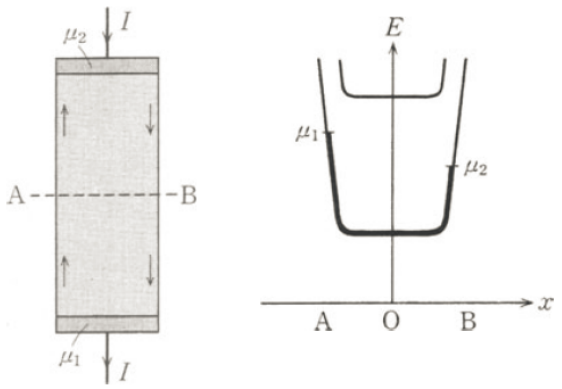

FIG. 8: Simple understanding based on Landauer picture.

defined by an infinitely high potential barrier. We consider a situation where the Fermi level is in between Landau levels, so the system is at the center of a quantum Hall plateau. The Fermi levels in the upper and lower electrodes may be different, for example the Fermi level in the lower electrode μ_1 may be higher than that in the upper electrode μ_2. In this case the edge states at the left edge are occupied up to /-Ll, but those at the right are occupied only up to μ_2. This nonequilibrium distribution of electrons will not be equilibrated in an ideal system, owing to the conservation of momentum.

Now we observe that the group velocity of each Landau level is

$$v_n = \frac{dE_n}{\hbar dk} \tag{71}$$

The current from each level n is canceled out between left and right edges and the only part that contributes is the difference contributes to the net current. So the current is given by

$$\begin{aligned} I_n &= e\int \frac{1}{2\pi} v_n dk && (72)\\ &= e\int \frac{1}{2\pi}\frac{dE_n}{\hbar dk} dk && (73)\\ &= e/h\int_{\mu_1}^{\mu_2} dE_n && (74)\\ &= e/h(\mu_2-\mu_1) = e^2/hV && (75) \end{aligned}$$

so that the total current is given by

$$I = (\#\text{occupied Landau levels})\frac{e^2}{h}V \tag{76}$$

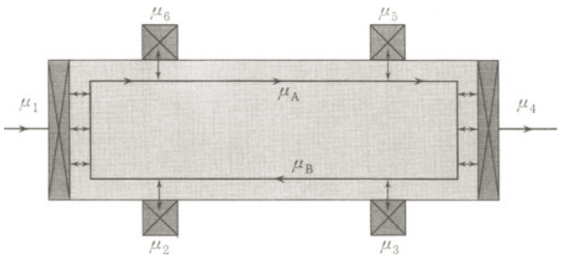

FIG. 9: Hall bar from the view point of edge states.

Six-terminal

With the above preparation, next we discuss the general case similar to the experimental setup (Fig. 9). At the upper edge, the two electrodes have the same chemical potential $\mu_5 = \mu_6 = \mu_A$. Likewise, at the lower edge, $\mu_2 = \mu_3 = \mu_B$, where $\mu_{A,B}$ are the chemical potentials of the upper and lower edges after equilibrium is attained at each edge.

Similarly, we assume that the left electrode has the chemical potental with the upper edge, and assume the electrons perfectly transport from electrode to the sample. So we have the current $I = Ne/h(\mu_1 - \mu_4) = Ne/h(\mu_A - \mu_B)$. The Hall voltage is $V_H = (\mu_A - \mu_B)/e$. So we get $I/V_H = Ne^2/h$, and $N = (\#\text{occupied Landau levels})$. Notice that it is self-consistent that, longitudinal resistance is zero: $R_L = (\mu_2 - \mu_3)/eI = (\mu_5 - \mu_6)/eI = 0$.

At last, we emphasize the astonishing feature of perfect transmission, which is independent of the length L (or more precisely of the aspect ratio L/W) or the particular geometry of the sample, may be understood from the edge-state picture which we have introduced above. Perfect transmission is due to that the electron cannot be backscattered unless it is scattered to the opposite edge with inverse chirality. However, in a usual quantum Hall system, the opposite edges are separated by a macroscopic distance W , and backscattering processes are therefore strongly (exponentially) suppressed, which determines the spatial extension of quantum-mechanical state, and the macroscopic sample width W .

THEORY OF IQHE: EXISTENCE OF PLATEAU, ROLE OF DISORDER

The theory of the IQHE, namely integral quantization of the Hall resistance, must answer the following questions: What is the origin of the quantization of the Hall resistance? What determines its value? Why is the quantization precise?

Let us take the classical formula for the Hall resistance

$$R_H = \frac{B}{\rho e}. \tag{77}$$

At the precise integer filling fraction, the longitudinal conductivity is zero, and the Hall conductivity is precisely the quantized value $R_H = \frac{B}{ne} = \frac{h}{ne^2}$, which is precisely the value of quantized Hall resistance. Is this the explanation of the Hall resistance? No, because it does not explain the plateaus. (In experiment, generically the chemical potential does sit in the degenerate band and the filling fraction is tuned continuously and is not quantized.)

While the system does have a gap under integer filling, we will need something that will preserve the special properties of the state even when we move away from the filling fraction which is precisely an integer. There happens to be a way out. An assumption made above, that of translational invariance, is not satisfied in the actual experimental system because of the unavoidable presence of disorder. Disorder, surprisingly, is crucial for the establishment of plateaus.

The effect of disorder

Tip: Anderson Localization. 1958 P. W. Anderson (for which he won the Nobel prize in 1977).

Disorder obviously modifies the spectrum. Much work has been done on how the single electron eigenstates are affected by the presence of disorder. The resulting picture is shown schematically in Fig. 10. Disorder will spread out the energies in the band by having some regions where the potential is higher than average and some regions where the potential is lower than average. This spreads the sharp peak in the density of states into a broadened band. (These ideas were developed by Iordansky , Kazarinov and Luryi 1982, Prange and Joynt 1982, and Trugman 1983)

To develop a feel for the effect of disorder on single particle eigenstates, let us consider a disorder potential that is smooth on the scale of the magnetic length, which will allow us

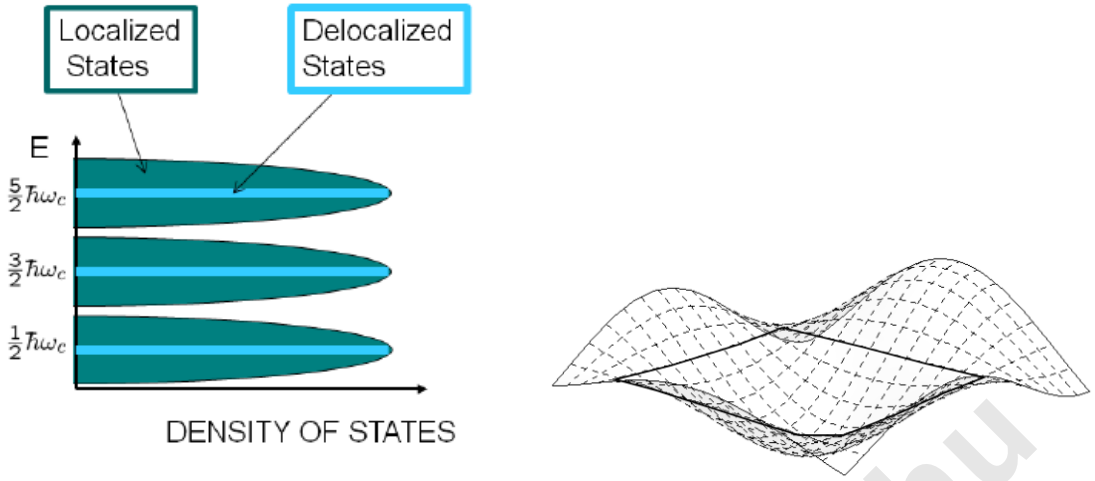

FIG. 10: The density of states for spin-polarized (or spinless) electrons in a magnetic field with disoder. The Landau bands are spread out, with localized eigenstates in the tails and extended eigenstates near the middle. Disorder potential configurations.

to use our semiclassical intuition. We imagine a potential landscape with hills and valleys. Landau levels are locally well defined for such a potential. For a given energy, an electron moves along an equipotential contour. Another way to see this is that the $E \times B$ drift and $E = -\partial V_{dis}(r)$. Thus, the disorder potential traps the electron into a closed orbit. This, interestingly, is true independent of whether the potential is attractive (a potential well) or repulsive (a potential hill).

Since current tends to flow perpendicular to potential gradients (i.e., it is hall current), eigenstates tend to follow contours of constant potential. Thus many of the eigenstates at high and low energy will be trapped in local minima or maxima isolated in a hill or valley and circling the peak or bottom. (i.e. If the equipotential lines are closed, as it is the case for most of the equipotential lines in a potential landscape, an electron cannot move from one point to another one over a macroscopic distance. An electron moving on a closed equipotential line can therefore not contribute to the electronic transport, and the electron is thus localised. Notice that this type of localisation it different from other popular types.) The result is that the eigenstates at the edge of the band experience localization, whereas (at least one) eigenstates near the center of the band are extended. When the chemical potential is anywhere in the localized states, then at low enough temperature, the electrons cannot move at all. Although there are states at this energy, they are all localized and

electrons cannot jump between them. Hence we expect in this case that the longitudinal conductance goes to zero (it is insulting).

The point that is not obvious, is why the Hall conductance should be precisely quantized. We will explain it in the following sections.

Exercises. Prove that, in the semiclassical picture, velocity satisfies "no source" condition, so that the disorder doesnot change the "conducting" electron.

$$\mathbf{E} + \mathbf{v} \times \mathbf{B} = 0 \tag{78}$$

$$(\mathbf{E} + \mathbf{v} \times \mathbf{B}) \times \frac{\mathbf{B}}{B^2} = 0 \tag{79}$$

$$-\partial V \times \frac{\mathbf{B}}{B^2} - \mathbf{v} = 0 \tag{80}$$

$$\mathbf{v} = -\partial V \times \frac{\mathbf{B}}{B^2} \tag{81}$$

$$\partial \cdot \mathbf{v} = \partial \cdot (\tfrac{\mathbf{B}}{B^2} \times \partial V) \tag{82}$$

$$= \partial_i (\tfrac{\mathbf{B}}{B^2} \times \partial V)_i \tag{83}$$

$$= \tfrac{1}{B} \partial_i \epsilon_{i3j} \partial_j V = 0 \tag{84}$$

Interpretation to experiment

We start with the situation of n completely filled LLs [column (a) of Fig. 11], which we have extensively discussed above: the Landau level n (and its potential landscape) is unoccupied. In a six-terminal measurement, one therefore measures the Hall resistance $R_H = h/e^2 n$ and a zero longitudinal resistance. Therefore, the quantum Hall system at integer filling factors is therefore a very unusual electron liquid: it is indeed a bulk insulator with perfectly conducting (non-dissipative) edges.

In column (b) of Fig. 11, we represent the situation where the Landau Level n gets moderately filled by electrons when the magnetic field B is decreased. These electrons in n populate preferentially the valleys of the potential landscape, or more precisely the closed equipotential lines that enclose these valleys. The electrons in the Landau Level n are thus (classically) localised somewhere in the bulk and do not affect the global transport characteristics, measured by the resistances, because they are not probed by the sample contacts. Therefore, the Hall resistance remains unaltered and the longitudinal resistance

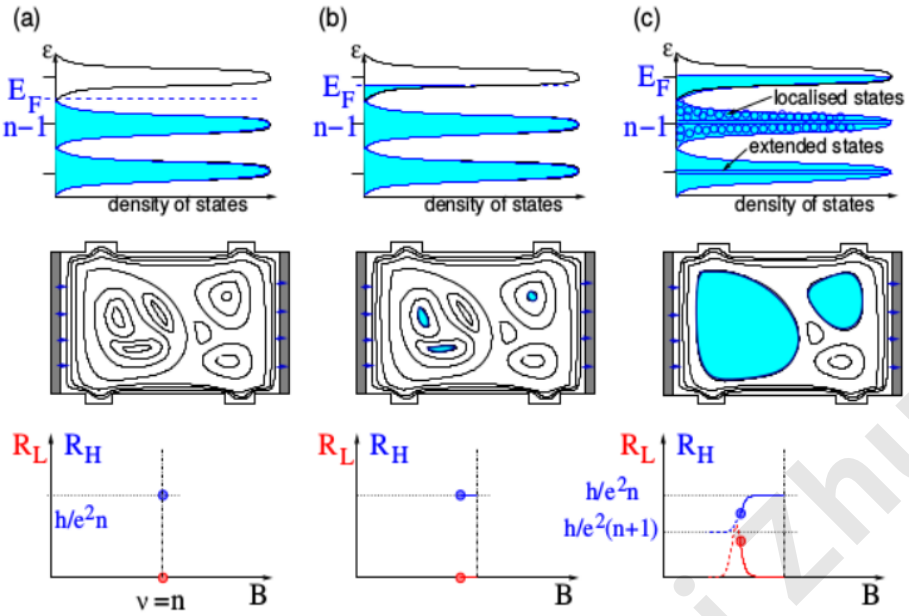

FIG. 11: The disorder-broadened density of states is shown in the first line for increasing fillings (a) - (c) described by the Fermi energy. The second line represents the disorder-potential landscape the valleys of which become successively filled with electrons when increasing the filling factor, i.e. when lowering the magnetic field at fixed particle number. The third line shows the corresponding Hall (blue) and the longitudinal (red) resistance measured in a six-terminal geometry, as a function of the magnetic field. The first figure in column (c) indicates that the bulk extended states are in the centre of the DOS peaks, whereas the localised states are in the tails.

remains zero despite the change of the magnetic field. This is the origin of the plateau in the Hall resistance.

If one continues to lower the magnetic field, the regions of the potential landscape in the Landau Level n occupied by electrons become larger, and they are eventually enclosed by equipotential lines that pass through the bulk and that connect the opposite edges. In this case, an electron injected at the left contact and travelling a certain distance at the upper edge may jump into the state associated with this equipotential line and thus reach the lower edge. Due to its chirality, the electron is then backscattered to the left contact, which causes an increase in the longitudinal resistance. Indeed, if one measures the resistance between the two contacts at the lower edge, a potential drop is caused by the electron that leaks in from this equipotential connecting the upper and the lower edge. It is this potential drop

that causes a non-zero longitudinal resistance. At the same moment the Hall resistance is no longer quantised and jumps to the next (lower) plateau, a situation that is called plateau transition. This situation of electron-filled equipotential lines connecting opposite edges, which are thus extended states [see first line of Fig. 11(c)] as opposed to the bulk localized states, arises when the Landau Level n is approximately half-filled.

Plateau transitions and scaling laws

The physical picture presented above suggests that the plateau transition in the Hall resistance is related to a percolation transition, where initially separated electron-filled valleys start to percolate between the opposite sample edges beyond a certain threshold of the filling. Because of the second-order character of a percolation transition, this scenario suggests that the plateau transition is a second-order quantum phase transition described by universal scaling laws, where the control parameter is just the magnetic field B.

The phase transition occurs at the critical magnetic field B_c and is characterised by an algebraically diverging correlation length

$$\xi \sim \frac{1}{|B - B_c|^{\nu}} \tag{85}$$

where ν is called the critical exponent.

THEORY OF IQHE: THOULESS PICTURE

Tips: The Kubo formula is a particular formulation of perturbation theory in quantum mechanics3 which turned out to be extremely useful in explaining the quantum Hall effect as first shown by Thouless (1982).

1. Framework

To the measure an observable operator $\hat{A}$ in the system described by the Hamiltonian $\mathcal{H}_0$, we now consider a perturbation $\hat{H}_1 = \hat{B}(t)f(t)$. The whole system is described by $H = H_0 + H_1$, and $f(t \to -\infty) = 0$ and $f(t \to 0) = \lambda << 1$. The response is

$$< A(t) > \; = \; < A > + \int_{-\infty}^{\infty} \chi_{AB}(t-t')f(t')dt' \tag{86}$$

$$\chi_{AB}(t-t') \; = \; i < [A(t), B(t')] >_0 \vartheta(t-t') \tag{87}$$

$\chi_{AB}(t-t')$ is real-time response function.

$$< A(\tau) > \; = \; < A > + \int_0^{\beta} \chi_{AB}(\tau-\tau')f(\tau')d\tau' \tag{88}$$

$$\chi_{AB}(\tau-\tau') \; = \; < T_\tau A(\tau)B(\tau')] >_0 \tag{89}$$

where $\chi_{AB}(\tau-\tau')$ is imagine-time response function.

$$< A(q,\omega) > \; = \; \chi_{AB}(q,\omega)f(\omega) \tag{90}$$

$$\chi_{AB}(q,\omega) = \int_{-\infty}^{\infty} e^{i\omega t}\chi_{AB}(t-t')\vartheta(t-t') = i\int_0^{\infty} e^{i\omega t} < [A(t), B(0)] >_0 \tag{91}$$

where $\chi_{AB}(q,\omega)$ is response function in frequency.

Proof.

$$\begin{aligned} A_H(t) \; &= \; e^{iHt}A_S e^{-iHt} = e^{i(H_0+H_1)t}A_S e^{-i(H_0+H_1)t} \\ &= \; e^{iHt}e^{-iH_0t}e^{iH_0t}A_S e^{-iH_0t}e^{iH_0t}e^{-i(H_0+H_1)t} \\ &= \; U^\dagger(t)A_I(t)U(t) \end{aligned} \tag{92}$$

where $U(t) = e^{iH_0t}e^{-i(H_0+H_1)t}$ and $A_I(t) = e^{iH_0t}A_S e^{-iH_0t}$.

And we know,

$$\begin{aligned} U(t) &= Texp\big[i\int_{-\infty}^{t} dt' B_I(t') f(t')\big] \\ &= 1+\sum_{n=1}\frac{(-i)^n}{n!}\int_0^t dt_1\int_0^t dt_2..\int_0^t dt_n T[B(t_1)B(t_2)...B(t_n)]f(t_1)... \\ &\approx 1+i\int_{-\infty}^{t} dt' B_I(t') f(t') \end{aligned} \tag{93}$$

So that,

$$\begin{aligned} A_H(t) &= [1-i\int_{-\infty}^{t} dt' B_I(t') f(t')] A_I(t) [1+i\int_{-\infty}^{t} dt' B_I(t') f(t')] \\ &\simeq A_I(t)+i\int_{-\infty}^{t} dt' [A_I(t), B_I(t')] f(t') \\ &= A_I(t)+\int_{-\infty}^{\infty} dt' \chi_{AB}(t-t') f(t') \end{aligned} \tag{94}$$

$$\chi_{AB}(t-t') = i[A_I(t), B_I(t')]\vartheta(t-t') \tag{95}$$

For single particle,

$$\begin{aligned} \chi_{AB} &= i\int_0^\infty dt e^{i\omega t} < [A(t), B(0)] > \\ &= i\int_0^\infty dt e^{i\omega t} < \sum_{m,k} A_{mk} a_m^\dagger a_k e^{-i(E_k-E_m)t/\hbar} \sum_{i,i} B_{ij} a_i^\dagger a_j - ... > \\ &= i\int_0^\infty dt e^{i\omega t} \sum_{m,k}\sum_{i,i} e^{-i(E_k-E_m)t/\hbar} A_{mk} B_{ij} (< a_m^\dagger a_k a_i^\dagger a_j > - < a_i^\dagger a_j a_m^\dagger a_k >) \\ &= i\int_0^\infty dt e^{i\omega t} \sum_{m,k}\sum_{i,i} e^{-i(E_k-E_m)t/\hbar} A_{mk} B_{ij} (n(E_m)-n(E_k))\delta_{mj}\delta_{ik} \\ &= -\sum_{m,k} A_{mk} B_{km} \frac{n(E_m)-n(E_k)}{\hbar\omega+i\eta-(E_k-E_m)} \end{aligned} \tag{96}$$

We have used the wick-theorem: $< a_m^\dagger a_k a_i^\dagger a_j >=< a_m^\dagger a_k >< a_i^\dagger a_j > + < a_m^\dagger a_j >< a_k a_i^\dagger > - < a_m^\dagger a_i^\dagger >< a_k a_j >$. And $< a_m^\dagger a_j >= \delta_{mj} n(E_m)$, $< a_k a_i^\dagger >= \delta_{ik}(1-n(E_k))$.

2. Several examples

Example 1. The electronic field $\vec{E} = -\nabla\varphi$, the perturbation to the system is $\rho\cdot\varphi$. So the change of density os the system is

$$\delta\rho(q,\omega) = e\chi_{nn}(q,\omega)\cdot\varphi(q,\omega) \tag{97}$$

where $\chi_{nn}(q,\omega) = i\int_0^\infty e^{i\omega t} < [\rho(t),\rho(0)] >_0$ and ρ is electronic density.

Example 2. The current induced by electronic field is $j_\alpha(q,\omega) = \sigma_{\alpha\beta}\cdot E_\beta$. With the help of $e\partial_t\rho + \nabla\cdot\rho = 0$

$$e\chi_{nn}(q,\omega)\varphi(q,\omega) = \delta\rho(q,\omega) = \frac{1}{\omega e}\vec{q}\cdot\vec{j}(q,\omega) = \frac{1}{\omega e}q_\alpha[-i\sigma_{\alpha\beta}(q,\omega)q_\beta\varphi(q,\omega)]$$

$$\Rightarrow \sigma_{\alpha\beta} = -\frac{ie^2}{q_\alpha q_\beta}\omega\chi_{nn}(q,\omega) \tag{98}$$

3. Conductivity

The perturbation from electronic field is

$$H_1 = \int d\mathbf{r} j_\alpha(\mathbf{r})A_\alpha(\mathbf{r},t) = \sum_{\mathbf{q}} j_\alpha(\mathbf{q},t)A_\alpha(-\mathbf{q},t) \tag{99}$$

here we choose the Coulomb gauge for electric field $-\partial_t A_\alpha(r,t) = E_\alpha(\mathbf{r},t)$. If this is not transparent, it is better to think about the energy in the real space, i.e. $eX_\alpha E_\alpha = e\sum_{i=1}^{N_e} x_{i\alpha}E_\alpha$. If we write $E_\alpha = -\partial_t A_\alpha(t)$, we will have $-e\sum_{i=1}^{N_e} x_{i\alpha}\partial_t A_\alpha(t) = -e\sum_i \partial_t x_{i\alpha}A_\alpha(t) + const. = -e\sum_i v_{i\alpha}A_\alpha(t) = j_\alpha A_\alpha(t)$.

If we further assume that the current and electric field is not uniform, we define the current operator as $j_\alpha(r) = e\sum_i[v_i^\alpha\delta(r-r_i) + \delta(r-r_i)v_i^\alpha]$ (Schrodinger picture).

If the electric field is taken $E_\alpha(r,t) = \Xi_\alpha\exp(i\vec{q}\cdot\vec{r} - i\omega t)$, we have $A_\alpha(r,t) = \frac{i}{\omega}\Xi_\alpha\exp(i\vec{q}\cdot\vec{r} - i\omega t)$, since $E_\alpha = -\partial_t A_\alpha(t)$. And $E_\alpha(q,t) = \int dr E_\alpha(r,t)e^{-iqr} = \Xi_\alpha e^{-i\omega t}$, $A_\alpha(q,t) = \frac{i}{\omega}\Xi_\alpha e^{-i\omega t}$.

The measured current is the averaged velocity of particles in the system, which is sum: $J_\alpha = \frac{e}{V}\sum_i v_i$, where $v_i = \frac{1}{m}(p_i - eA(r_i,t))$. So that $J_\alpha = \frac{e}{V}\sum_i < p_i > -\frac{e^2}{mV}\sum_i < A(r_i,t) >=< j_\alpha(r,t) > + i\frac{n_0e^2}{m\omega}E_\alpha(r,t) = J^{(1)} + J^{(2)}$. $J^{(1)}$ is a constant part and not contributed to real part of conductivity.

$J^{(2)}$ can be obtained from linear response theory:

$$\begin{aligned}
J^{(2)} =< j_\alpha(r,t) > &= < \psi'|e^{iHt}j_\alpha(r)e^{-iHt}|\psi' > \\
&= -i\int_{-\infty}^{t} dt' < [j_\alpha(r,t), H_1'(t')] > \\
&= \frac{1}{\omega}\Xi_\beta\int_{-\infty}^{t} dt' e^{-i\omega t'} < [j_\alpha(r,t), j_\beta(q,t')] > \\
&= \frac{1}{\omega}E_\beta(r,t)e^{-iqr}e^{i\omega t}\int_{-\infty}^{t} dt' e^{-i\omega t'} < [j_\alpha(r,t), j_\beta(q,t')] > \\
&= \sigma_{\alpha\beta}E_\beta(r,t)
\end{aligned} \tag{100}$$

Noted that we have used the interaction picture:

$$\begin{aligned} j_\alpha(r,t) &= e^{iH_0t} j_\alpha(r) e^{-iH_0t} \\ H_1'(t) &= e^{iH_0t} H_1' e^{-iH_0t} \end{aligned}$$

So we obtain,

$$\begin{aligned} \sigma_{\alpha\beta}(r,t) &= < J^{(2)}(r,t) > / E_\beta(r,t) \\ &= \frac{1}{\omega} e^{-iqr} \int_{-\infty}^{t} e^{i\omega(t-t')} < [j_\alpha^\dagger(r,t), j_\beta(q,t')] > \\ \text{(To average in space)} \Rightarrow & \frac{1}{\omega V} \int_{-\infty}^{t} e^{i\omega(t-t')} < [j_\alpha^\dagger(q,t), j_\beta(q,t')] > \end{aligned} \tag{101}$$

We have used $\int d^3r e^{-iq\cdot r} j_\alpha(r,t) = j_\alpha(-q,t) = j_\alpha^\dagger(q,t)$.

The retarded correction function:

$$\Pi_{\alpha\beta}(t-t',q) = -\frac{i}{V}\vartheta(t-t') < [j_\alpha^\dagger(q,t), j_\beta(q,t')] > \tag{102}$$

$$\Pi_{\alpha\beta}(\omega,q) = -\frac{i}{V}\int_{-\infty}^{\infty} dt e^{i\omega(t-t')}\vartheta(t-t') < [j_\alpha^\dagger(q,t), j_\beta(q,t')] > \tag{103}$$

$$\sigma_{\alpha\beta} = \frac{i}{\omega}\Pi_{\alpha\beta}(\omega,q) \tag{104}$$

$$Re[\sigma_{\alpha\beta}] = -\lim_{\omega\mapsto 0}\frac{Im\Pi_{\alpha\beta}(\omega,q)}{\omega} \tag{105}$$

With the help of Eq. 96

$$\begin{aligned} Re[\sigma_{\alpha\beta}](\omega) &= -\frac{1}{\omega V} Im[\sum_{q,k} < q|j_\alpha|k >< k|j_\beta|q > \frac{n_F(E_q) - n_F(E_k)}{\hbar\omega + i\eta - (E_k - E_q)}] \\ &= \frac{1}{\omega V}\sum_{q,k} < q|j_\alpha|k >< k|j_\beta|q > [n_F(E_q) - n_F(E_k)]\pi\delta(\hbar\omega - (E_k - E_q)) \\ &= \frac{1}{\omega V}\int_{-\infty}^{\infty} dE \sum_{q,k} < q|j_\alpha|k >< k|j_\beta|q > \pi\delta(E + \hbar\omega - E_k)\delta(E - E_q) \\ &= \frac{1}{\pi V}\int_{-\infty}^{\infty} dE \frac{n_F(E) - n_F(E+\hbar\omega)}{\omega} \\ &\quad \sum_{q,k} < q|j_\alpha|k >< k|j_\beta|q >< q|ImG^R(E)|q >< k|ImG^R(E+\hbar\omega)|k > \\ &= \frac{\hbar}{\pi V}\int_{-\infty}^{\infty} dE \frac{n_F(E) - n_F(E+\hbar\omega)}{\hbar\omega} Tr[j_\alpha ImG^R(E) j_\beta ImG^R(E+\hbar\omega)] \end{aligned} \tag{106}$$

If we consider the dc conductivity $\hbar\omega \mapsto 0$ and zero temperature,

$$\sigma_{\alpha\beta}(E_F) = \frac{\hbar}{\pi V} Tr[j_\alpha ImG^R(E_F) j_\beta ImG^R(E_F)] \tag{107}$$

Alternatively, we need another formula, let us start from Eq. again:

$$\begin{aligned} Re[\sigma_{\alpha\beta}](\omega) &= -\frac{1}{\omega V} Im[\sum_{q,k} n_F(E_q) < q|j_\alpha|k >< k|j_\beta|q > \frac{1}{\hbar\omega + i\eta - (E_k - E_q)} \\ &\quad + n_F(E_q) < k|j_\alpha|q >< q|j_\beta|k > \frac{1}{-\hbar\omega - i\eta + (E_q - E_k)}] \\ &= \sigma^1 + \sigma^2 \end{aligned} \tag{108}$$

$$\begin{aligned} \sigma^1 &= -\frac{1}{\omega V} Im[\sum_{q,k} n_F(E_q) \frac{< q|j_\alpha|k >< k|j_\beta|q > + < k|j_\alpha|q >< q|j_\beta|k >}{E_k - E_q}] \\ &= -\frac{e^2}{\omega V} Im[\sum_{q,k} n_F(E_q) \frac{(E_q - E_k) < q|x_\alpha|k >< k|v_\beta|q > + (E_k - E_q) < k|x_\alpha|q >< q|v_\beta|k >}{E_k - E_q} \\ &= -\frac{e^2}{\omega V} Im[\sum_{q,k} n_F(E_q)[< q|x_\alpha|k >< k|v_\beta|q > - < k|x_\alpha|q >< q|v_\beta|k >] \\ &= -\frac{e^2}{\omega V} Im[\sum_{q} n_F(E_q) < q|[x_\alpha v_\beta - v_b etax_\alpha]|q > \\ &= 0, if \alpha \neq \beta \end{aligned} \tag{109}$$

Here we used $\frac{1}{\pm\hbar\omega + E_q - E_k} \approx \frac{1}{E_q - E_k}(1 \pm \frac{-\hbar\omega}{E_q - E_k})$.

$$\begin{aligned} \sigma^2 &= -\frac{1}{\omega V} Im[\sum_{q,k} n_F(E_q) \frac{[- < q|j_\alpha|k >< k|j_\beta|q > + < k|j_\alpha|q >< q|j_\beta|k >]\hbar\omega}{(E_k - E_q)^2}] \\ &= \frac{\hbar e^2}{V} Im[\sum_{q,k} n_F(E_q) \frac{< q|v_\alpha|k >< k|v_\beta|q > - < k|v_\alpha|q >< q|v_\beta|k >}{(E_k - E_q)^2}] \\ &= \frac{\hbar e^2}{V} Im[\sum_{q,k} (n_F(E_q) - n_F(E_k) \frac{< q|v_\alpha|k >< k|v_\beta|q >}{(E_k - E_q)^2}] \\ (T = 0) \rightarrow & \frac{\hbar e^2}{V} Im[\sum_{E_q < E_F < E_k} \frac{< q|v_\alpha|k >< k|v_\beta|q > - < k|v_\alpha|q >< q|v_\beta|k >}{(E_k - E_q)^2}] \end{aligned} \tag{110}$$

This the Hall conductivity written via the Kubo formula.

Then Qian Niu proposed that the wave function for the opposite edges are related by magnetic translations

$$\psi(x + a, y) = e^{i\theta_x a} e^{i\frac{a}{l^2}y} \psi(x, y) \tag{111}$$

$$\psi(x, y + b) = e^{i\theta_y b} \psi(x, y) \tag{112}$$

In previous note, the single-particle wave function is

$$\psi_{N,j}(x,y) = \left(\frac{1}{2^N N! \pi^{1/2} bl}\right)^{1/2} \sum_{k=-\infty}^{\infty} \exp[i\frac{X_j + ka}{l^2}y - \frac{(X_j + ka - x)^2}{2l^2}]H_N(\frac{X_j + ka - x}{l}) \tag{113}$$

which satisfies the boundary condition $|\vec{L}_1| = a, |\vec{L}_2| = b$:

$$\psi(x + a, y) = e^{i\frac{a}{l^2}y}\psi(x, y) \tag{114}$$

$$\psi(x, y + b) = \psi(x, y) \tag{115}$$

Next we use a different definition for the single particle wave function:

$$\psi_{N,m}(x,y) = \left(\frac{1}{2^N N! \pi^{1/2} bl}\right)^{1/2} \sum_{n=-\infty}^{\infty} e^{ik_x X_n} \exp[ik_y y - \frac{(k_y l^2 - x)^2}{2l^2}]H_N(\frac{k_y l^2 - x}{l}) \tag{116}$$

where the effective momentum is

$$k_x = \frac{2\pi m}{a}, k_y = \frac{X_n}{l^2} = \frac{2\pi n}{b} \tag{117}$$

One can check that the boundary condition is:

$$\psi(x + a, y) = e^{i\frac{a}{l^2}y}\psi(x, y) \tag{118}$$

$$\psi(x, y + b) = \psi(x, y) \tag{119}$$

That means Eq. 116 is equivalent to the original definition.

Then we introduce twisted boundary condition (θ_1, θ_2):

$$k_x^\theta = \theta_x + \frac{2\pi m}{a}, k_y^\theta = \theta_y + \frac{2\pi n}{b} \tag{120}$$

and single-particle wave function becomes:

$$\psi_{N,m}^\theta(x,y) = \left(\frac{1}{2^N N! \pi^{1/2} bl}\right)^{1/2} \sum_{n=-\infty}^{\infty} e^{ik_x^\theta X_n} \exp[ik_y^\theta y - \frac{(k_y^\theta l^2 - x)^2}{2l^2}]H_N(\frac{k_y^\theta l^2 - x}{l}) \tag{121}$$

We get that

$$\psi(x + a, y) = e^{i\theta_x a}e^{i\frac{a}{l^2}y}\psi(x, y) \tag{122}$$

$$\psi(x, y + b) = e^{i\theta_y b}\psi(x, y) \tag{123}$$

which is the required boundary condition under (θ_x, θ_y): adding a phase factor on boundary condition (comparing Eq. 97,98 and 101,102).

Now we make the unitary transformation on the wave function

$$\psi(x,y) \to e^{-i\theta_x x - i\theta_y y}\psi(x,y) \tag{124}$$

, which is equivalent to change hamiltonian to

$$\begin{aligned} &H(x,y) \to \overline{H}, \\ &\overline{H} = H(-i\partial_x + \theta_x, -i\partial_y + \theta_y) \end{aligned} \tag{125}$$

, similar to shift the momentum (twisted boundary condition). And we have the velocity can be expressed as $v_{x,y} = \frac{\partial \overline{H}}{\hbar \partial \theta_{x,y}}$.

Then we have

$$\begin{aligned} \sigma_H &= \frac{\hbar e^2}{V} Im[\sum_{q,k} n_F(E_q) \frac{< q|v_\alpha|k >< k|v_\beta|q > - < k|v_\alpha|q >< q|v_\beta|k >}{(E_k - E_q)^2}] \\ &= \frac{\hbar e^2}{\hbar^2 V} Im[\sum_{q,k} n_F(E_q) \frac{< q|\frac{\partial \overline{H}}{\partial \theta_x}|k >< k|\frac{\partial \overline{H}}{\partial \theta_y}|q > - < k|\frac{\partial \overline{H}}{\partial \theta_y}|q >< q|\frac{\partial \overline{H}}{\partial \theta_x}|k >}{(E_k - E_q)^2}] \\ &= \frac{\hbar e^2}{\hbar^2 V} Im[\sum_{E_q < E_F} < \frac{\partial q}{\partial \theta_x}|\frac{\partial q}{\partial \theta_y} > - < \frac{\partial q}{\partial \theta_y}|\frac{\partial q}{\partial \theta_x} >] \end{aligned} \tag{126}$$

Here we used the relation $\langle n'|\nabla n\rangle = \frac{\langle n'|\nabla H|n\rangle}{E_n - E_{n'}}$.

One major condition for quantization in this approach is that the Hall conductivity is a local response function, insensitive to the boundary condition. We can therefore average over all the phases $\theta \in [0, 2\pi]$ (in fact, we made a transformation $\theta'_{x(y)} = \theta_{x(y)} a(b)$, so that the volume can be cancelled.) that specify different boundary conditions

$$\begin{aligned} \sigma_H &= \frac{e^2}{h} \int_0^{2\pi} \int_0^{2\pi} d\theta_x d\theta_y \frac{1}{2\pi i} [< \frac{\partial q}{\partial \theta_x}|\frac{\partial q}{\partial \theta_y} > - < \frac{\partial q}{\partial \theta_y}|\frac{\partial q}{\partial \theta_x} >] \\ &= \frac{e^2}{h} \int_0^{2\pi} \int_0^{2\pi} d\theta_x d\theta_y \frac{1}{2\pi i} \partial \times \mathbf{A} \\ &= \frac{e^2}{h} \oint dl \cdot \mathbf{A} \frac{1}{2\pi i} \\ &= \frac{e^2}{h} N \end{aligned} \tag{127}$$

The berry connection is defined as

$$A_\alpha = < \frac{\partial q}{\partial \theta_\alpha}|q > - < q|\frac{\partial q}{\partial \theta_\alpha} > \tag{128}$$

4. Berry phase

Assuming a physical system is depended on some parameters $\mathbf{R} = (R_1, R_2, \cdots, R_N)$, we have the snapshot Hamiltonian $H(\mathbf{R})$, its eigen-values and eigen-states:

$$H(\mathbf{R})|n(\mathbf{R})\rangle = E_n(\mathbf{R})|n(\mathbf{R})\rangle \tag{129}$$

where $|n(\mathbf{R})\rangle$ can have an arbitrary phase prefactor.

The parameters $\mathbf{R}(\mathbf{t})$ are slowly changed with time t, then the *adiabatic* evolution of time-dependent Schrodinger equation:

$$i\frac{d}{dt}|\psi(t)\rangle = H(\mathbf{R}(\mathbf{t}))|\psi(t)\rangle \tag{130}$$

Take the Ansatz $|\psi(t)\rangle = e^{i\gamma_n(t)}e^{-i\int_0^t E_n(\mathbf{R}(\mathbf{t}'))dt'}|n(\mathbf{R}(\mathbf{t}))\rangle$, we have

$$i\frac{d}{dt}|\psi(t)\rangle = e^{i\gamma_n(t)}e^{-i\int_0^t E_n(\mathbf{R}(\mathbf{t}'))dt'}\left[-\frac{d}{dt}\gamma_n(t) + E_n(R(t)) + |\frac{d}{dt}n(\mathbf{R}(t))\rangle\right] \tag{131}$$

Here $e^{-i\int_0^t E_n(\mathbf{R}(\mathbf{t}'))dt'}$ is the dynamic phase, and γ_n is the geometric phase which will be clarified below.

We insert the above Ansatz into the rhs of the Schrdinger equation and use the "adiabatic" condition $H(R(t))|\psi(t)\rangle = E_n(R(t))|\psi(t)\rangle$,

$$-\left(\frac{d}{dt}\gamma_n\right)|n\rangle + i\left|\frac{d}{dt}n\right\rangle = 0 \tag{132}$$

Multiply from the left by $< n(R(t))|$, and obtain the Berry phase expression:

$$\frac{d}{dt}\gamma_n = i\langle n(R(t))|\frac{d}{dt}|n(R(t))\rangle = i\frac{R(t)}{dt}\langle n(R)|\bigtriangledown_R|n(R)\rangle \tag{133}$$

$$\gamma_n(\mathcal{C}) = \int_{\mathcal{C}} i\langle n(\mathbf{R})|\nabla_{\mathbf{R}}n(\mathbf{R})\rangle d\mathbf{R} \tag{134}$$

We can define Berry connection:

$$\mathbf{A}^{(n)}(\mathbf{R}) = i\langle n(\mathbf{R})|\nabla_{\mathbf{R}}n(\mathbf{R})\rangle = -Im\langle n(\mathbf{R})|\nabla_{\mathbf{R}}n(\mathbf{R})\rangle \tag{135}$$

Gauge transformation $|n(\mathbf{R})\rangle \to e^{i\alpha(\mathbf{R})}|n(\mathbf{R})\rangle$, $\mathbf{A}^{(n)}(\mathbf{R}) \to \mathbf{A}^{(n)}(\mathbf{R}) - \nabla_{\mathbf{R}}\alpha(\mathbf{R})$ and $\gamma = \oint \mathbf{A}(\mathbf{R})d\mathbf{R}$ is gauge invariant.

Define the Berry curvature:

$$\mathbf{B}(\mathbf{R}) = \nabla_{\mathbf{R}} \times \mathbf{A}^{(n)}(\mathbf{R}) \tag{136}$$

Using Stokes theorem, we have for the Berry Phase:

$$\gamma_n(\mathcal{C}) = \int_{\mathcal{S}} \mathbf{B}^{(n)}(\mathbf{R}) d\mathcal{S} \tag{137}$$

where $\mathcal{S}$ is any surface whose boundary is the loop $\mathcal{C}$. Alternatively, we can get the Berry phase using Berry curvature by using two useful formula:

$$B_j = \epsilon_{jkl}\partial_k A_l = -Im\epsilon_{jkl}\partial_k\langle n|\partial_l n\rangle = -Im\epsilon_{jkl}\langle\partial_k n|\partial_l n\rangle, \tag{138}$$

$$\mathbf{B}^{(n)} = -Im\sum_{n'\neq n}\langle\nabla n|n'\rangle\times\langle n'|\nabla n\rangle. \tag{139}$$

Moreover, we have

$$H(\mathbf{R})|n\rangle = E_n|n\rangle \tag{140}$$

$$\Rightarrow (\nabla H)|n\rangle + H|\nabla n\rangle = (\nabla E_n)|n\rangle + E_n|\nabla n\rangle \tag{141}$$

$$\Rightarrow \langle n'|\nabla H|n\rangle + \langle n'|H|\nabla n\rangle = 0 + E_n\langle n'|\nabla n\rangle \tag{142}$$

$$\Rightarrow \langle n'|\nabla n\rangle = \frac{\langle n'|\nabla H|n\rangle}{E_n - E_{n'}} \tag{143}$$

then we get:

$$\mathbf{B}^{(n)} = -Im\sum_{n'\neq n}\langle\nabla n|n'\rangle\times\langle n'|\nabla n\rangle = -Im\sum_{n'\neq n}\frac{\langle n|\nabla H|n'\rangle\times\langle n'|\nabla H|n\rangle}{(E_n - E_{n'})^2} \tag{144}$$

which recover the linear response theory!

Remark: Compare Berry phase with A-B phase $\gamma_n = \oint \mathbf{A}\cdot d\mathbf{R}(t)$ v.s. $\oint \mathbf{A}\cdot d\mathbf{r}$: Berry phase is a dynamical phase that is resulting from the adiabatical evolution. A-B phase is accumulating phase of a charged particle in the electric and magnetic field. In A-B phase, magnetic field or electric field is static, and the particle is moving.

Remark: Berry phase depends on the geometry of closed path.

Remark: To explicitly see the gauge invariance of the Berry curvature, we now notice that the eigenstates we have chosen previously have one point where they are not well defined. This will be extremely important in the Chern insulator problem: if we are able to find a gauge in which all wavefunctions are well defined, then the system cannot have nonzero Hall conductance.

One Example

Example: spin-1/2 in the presence of time-dependent magnetic field

$$H = -\mathbf{B}(t)\cdot\mathbf{S} = -[\sigma_x X(t) + \sigma_y Y(t) + \sigma_z Z(t)] \tag{145}$$

We will let the direction of B in space be the control parameter of the Hamiltonian: $\mathbf{R}(t) = (X(t), Y(t), Z(t)), R = |\mathbf{R}| = \sqrt{X^2+Y^2+Z^2}$.

The instantaneous eigenvalue is $|-\sigma\cdot\mathbf{R}(t) - E| = 0$, thus $E_m = mR(t), m = \pm$. The Berry phase is calculated below. First, we have $\nabla_{R(t)}H(R) = -\sigma_x\hat{e}_x - \sigma_y\hat{e}_y - \sigma_z\hat{e}_z$. Second, we calculate the element by choosing the direction of R along z-direction. $\sigma_z|\pm\rangle = \pm|\pm\rangle$, $\sigma_x|\pm\rangle = |-+\rangle$, $\sigma_y|\pm\rangle = \pm i|-+\rangle$. Third, $\langle+|\nabla_R H|-\rangle = \hat{e}_x + i\hat{e}_y$, and $\langle-|\nabla_R H|+\rangle = \hat{e}_x - i\hat{e}_y$. $\langle+|\nabla_R H|-\rangle \times \langle-|\nabla_R H|+\rangle = -2i\hat{e}_z$. Thus we have $\mathbf{B}^{(+)} = -Im\sum_{n'\neq n}\frac{\langle n|\nabla H|n'\rangle\times\langle n'|\nabla H|n\rangle}{(E_n - E_{n'})^2} = \frac{\hat{e}_z}{2R^2}$, and $\mathbf{B}^{-} = -\frac{\hat{e}_z}{2R^2}$. In general, for arbitrary direction $\vec{R}$, we have

$$\mathbf{B}^{(+)} = \frac{\vec{R}}{2R^3}, \mathbf{B}^{(-)} = -\frac{\vec{R}}{2R^3} \tag{146}$$

The resulting berry phase for a circle-like loop on the sphere is

$$\gamma_n = \int_{\mathcal{C}} \mathbf{B}^{(\pm)}\cdot dS = \pm\int_0^{2\pi} d\phi \int_0^{\theta} d\theta \sin\theta R\vec{R}\frac{\vec{R}}{2R^3} = \pm\pi(1-\cos\theta) = \pm\frac{\Omega(C)}{2} \tag{147}$$

where Ω is the solid angular enclosed by the closed path $\mathcal{C}$. Since the berry phase is related the geometric path, we also call it as geometric phase.

Note that $\mathbf{B}^{(+)} = \frac{\vec{R}}{2R^3}, \mathbf{B}^{(-)} = -\frac{\vec{R}}{2R^3}$ implies the magnetic monopole! This is the same formula we obtained earlier from the eigenstate formalism and is the flux through an area bounded by the curve C of a monopole with strength $\pm 1/2$ located at the degeneracy. The Berry curvature, which is gauge invariant, is singular at he origin $R = 0$ (but regular everywhere else).

If a monopole of strength e_M exists,, i.e. $\partial\cdot\mathbf{B} = 4\pi e_M\delta(R)$, then the magnetic field around it is $\mathbf{B} = e_M\frac{\mathbf{R}}{R^3}$, similar as the electric field generated by a charge. But electric field is related to the potential as $E = -\partial V$, it is impossible to find a regular-everywhere vector potienal that $\mathbf{B} = \partial\times\mathbf{A}$. To proof this, let us consider a sphere S^2 of radium $R = 1$ enclosing the monopole, and denote $\Sigma_{N(S)}$ as the north and south hemisphere, which meet at the equator C. Next we have $\int_{\Sigma_N}\mathbf{B}\cdot\mathbf{n}dS = \int_C \mathbf{A}\cdot dl$, and $\int_{\Sigma_S}\mathbf{B}\cdot\mathbf{n}dS = \int_{-C}\mathbf{A}\cdot dl$.

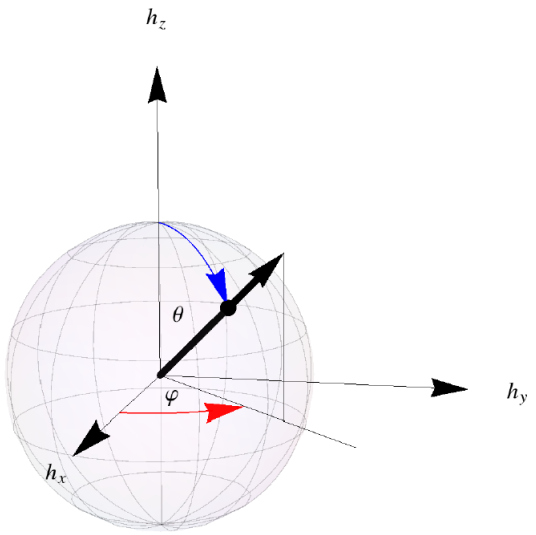

FIG. 12: Bloch sphere.

Summing together the two fluxes we conclude that $\int_{S^2} \mathbf{B} \cdot \mathbf{n} dS = \int_C \mathbf{A} \cdot dl + \int_{-C} \mathbf{A} \cdot dl = 0$. It is contrary to the fact that the flux should be $4\pi e_M$.

Again, another way of appreciating the singularity of vector potential, we can consider a small circle C, of radius $R\sin\theta$, encircling the north pole of spehere at an angle θ. We already know the magnetic flux through the solid angle $2\pi(1-\cos\theta)$ enclosed by C is $e_M 2\pi(1-\cos\theta)$. We can represent such a flux by the line integral of a vector potential $\mathbf{A}$ as

$$\int \mathbf{A} \cdot dl = A_\phi 2\pi R \sin\theta = e_M 2\pi(1-\cos\theta), \rightarrow A_\phi = \frac{e_M}{R}\frac{1-\cos\theta}{\sin\theta} \tag{148}$$

which becomes singlular at θ.

Another Example

The generic formof the Hamiltonian of any two level system is

$$H = \varepsilon(R)\sigma_0 + \mathbf{d}(R) \cdot \sigma \tag{149}$$

where ri are the Pauli matrices and d is a 3-D vector that depends on the coordinates R. This Hamiltonian describesmany interesting systems in condensed matter such as graphene, spinorbit- coupled systems, Bogoliubov quasiparticles, a spin-1/2 electron in a magnetic field, and many others. The energy levels are $E_\pm = \varepsilon(R) \pm \sqrt{\mathbf{d} \cdot \mathbf{d}}$. If we employ spherical

coordinates and parametrize the vector $d(R) = |d|(\sin(\theta)\cos(\phi), \sin(\theta)\sin(\phi), \cos(\theta))$. The two eigenstates with energies $\pm\sqrt{|d|}$ are

$$|R,-\rangle = \begin{pmatrix} \sin\frac{\theta}{2}e^{i\phi} \\ -\cos\frac{\theta}{2} \end{pmatrix}, |R,+\rangle = \begin{pmatrix} \cos\frac{\theta}{2}e^{-i\phi} \\ \sin\frac{\theta}{2} \end{pmatrix} \tag{150}$$

The two leftover Berry vector potential, A_θ and A_ϕ and the Berry curvature, $B_{\theta\phi}$, are given by

$$\mathbf{A} = A_R\vec{e}_r + A_\theta\vec{e}_\theta + A_\phi\vec{e}_\phi \tag{151}$$

$$A_r = i\langle R,-|\partial_R|R,-\rangle = 0 \tag{152}$$

$$A_\theta = \frac{1}{R}i\langle R,-|\partial_\theta|R,-\rangle = 0, \tag{153}$$

$$A_\phi = \frac{1}{R\sin\theta}i\langle R,-|\partial_\phi|R,-\rangle = -\frac{\sin^2\frac{\theta}{2}}{R\sin\theta} \tag{154}$$

$$B_{\theta\phi} = \vec{e}_r\frac{1}{R\sin\theta}[\partial_\theta(\sin\theta A_\phi) - \partial_\phi A_\theta] + \vec{\theta}\frac{1}{R}[\frac{1}{\sin\theta}\partial_\phi A_R - \partial_R(RA_\phi)] + \vec{e}_\phi\frac{1}{R}[\partial_R(RA_\theta) - \partial_\theta A_R] = -\frac{1}{2}\frac{\mathbf{R}}{R^3} \tag{155}$$

$$\gamma_n = \int \mathbf{B}_{\theta\phi}\cdot dS = \int_0^{2\pi} d\phi \int_0^{\pi} d\theta \sin\theta\frac{1}{2} = 2\pi \tag{156}$$

This is the Chern number (mod 2π) (we will clarify it later).

Please note that, the wavefunction $|R,-\rangle$ is not well defined if the system can reach, in its adiabatic evolution, the south pole $\theta = \pi$. In the case of a model of Chern insulator, we will see that we cannot pick a gauge that is everywhere well definedthe sphere will be fully covered.

The Berry phase in Bloch bands

The geometric phase is a critical concept in the modern theory of crystalline systems. Here we introduce the basic ideas that will enable us to apply the generic Berry phase formalism to band theory.

The electronic properties of the crystal are described within the independent particle approximation by the single-particle Hamiltonian $H = -\frac{\nabla^2}{2m} + V(r)$ where m is the electron mass and $V(r) = V(r+a)$ is a periodic potential. In the Brillouin zone (BZ), the eigenfunction takes the form of $\psi_{n,\mathbf{q}}(r) = e^{i\mathbf{q}\cdot\mathbf{r}}u_{n,\mathbf{q}}(\mathbf{r})$, according to the Bloch's theorem. We perform

a unitary transformation and introduce the cell-periodic function $u_{n,\mathbf{q}}(r) = e^{-i\mathbf{q}\cdot r}\psi_{n,\mathbf{q}}(\mathbf{r})$ and $u_{n,\mathbf{q}}(r) = u_{n,\mathbf{q}}(r+a)$. The price for this is a q-dependent Hamiltonian, $H(\mathbf{q}) = e^{-i\mathbf{q}\cdot r}He^{i\mathbf{q}\cdot r}$:

$$H(\mathbf{q}) = -\frac{(\nabla + i\mathbf{q})^2}{2m} + V(r) \tag{157}$$

which can be obtained by Schrodinger equation $H\psi_{n,\mathbf{q}} = E(\mathbf{q})\psi_{n,\mathbf{q}}$ and $e^{-i\mathbf{q}\cdot r}He^{i\mathbf{q}\cdot r}u_{n,\mathbf{q}} = E(\mathbf{q})u_{n,\mathbf{q}}$

At this point, we can make the connection with the general Berry phase formalism. We have a parameter dependent Hamiltonian, $H(R) \to H(q)$, and a single Hilbert space $\mathcal{H}$, so that for each value of the parameter $R \to q$ the set $|n;R\rangle \to u_{n;q}$ represents an orthonormal basis for H. The parameter space is the Brillouin zone, $M \to BZ$. A slow cyclic variation of q in the BZ, which may by caused by an external field that enters the Schrodinger equation as a time-dependent variation of the wave vector, $q \to q + q(t)$, will result in the wave function acquiring a geometric phase.

$$\begin{aligned}
\langle q|v_\gamma|q\rangle &= \int dr\psi^*_{n,\mathbf{q}}[-\frac{(i\nabla - e\mathbf{A})_\gamma}{m}]\psi_{n,\mathbf{q}} \\
&= \int dru^*_{n,\mathbf{q}}(r)e^{-i\mathbf{q}\cdot r}[-\frac{(i\nabla - eA)_\gamma}{m}]u_{n,\mathbf{q}}(r)e^{i\mathbf{q}\cdot r} \\
&= \int dru^*_{n,\mathbf{q}}(r)[-\frac{(i\nabla + \mathbf{q} - e\mathbf{A})_\gamma}{m}]u_{n,\mathbf{q}}(r) \\
&= \int dru^*_{n,\mathbf{q}}(r)\frac{\partial H(\mathbf{q})}{\partial \mathbf{q}}u_{n,\mathbf{q}}(r) \\
&= \langle u_{n,q}|\frac{\partial H(\mathbf{q})}{\partial \mathbf{q}}|u_{n,q}\rangle
\end{aligned}$$

Thus the Hall conductance becomes

$$\begin{aligned}
\sigma_H &= \frac{\hbar e^2}{V}Im[\sum_{q,k} n_F(E_q)\frac{<q|v_\alpha|k><k|v_\beta|q> - <k|v_\alpha|q><q|v_\beta|k>}{(E_k - E_q)^2}] \\
&= \frac{\hbar e^2}{\hbar^2 V}Im[\sum_{q,k} n_F(E_q)\frac{<q|\frac{\partial \overline{H}}{\partial k_x}|k><k|\frac{\partial \overline{H}}{\partial k_y}|q> - <k|\frac{\partial \overline{H}}{\partial k_y}|q><q|\frac{\partial \overline{H}}{\partial k_x}|k>}{(E_k - E_q)^2}] \\
&= \frac{\hbar e^2}{\hbar^2 V}Im[\sum_{E_q<E_F} <\frac{\partial q}{\partial k_x}|\frac{\partial q}{\partial k_y}> - <\frac{\partial q}{\partial k_y}|\frac{\partial q}{\partial k_x}>]
\end{aligned} \tag{158}$$

To account for these effects, we introduce the Berry connection and the corresponding

Berry curvature vector

$$\mathbf{A}^n(\mathbf{q}) = i\langle u_n(\mathbf{q})|\nabla_{\mathbf{q}}|u_n(\mathbf{q})\rangle \tag{159}$$

$$\mathbf{B}^n = \nabla \times \mathbf{A}^n(\mathbf{q}) = i\langle \nabla_{\mathbf{q}} u_n(\mathbf{q})| \times |\nabla_{\mathbf{q}} u_n(\mathbf{q})\rangle \tag{160}$$

Similarly, the berry phase is defined in integral in the BZ: $\gamma_n = \int_{C_{BZ}} d\mathbf{q} \cdot i\langle u_n(\mathbf{q})|\nabla_{\mathbf{q}}|u_n(\mathbf{q})\rangle$.

The winding number

In the previous sections we have proved that the Hall conductance of a filled band is equal to the integral of the Berry curvature over the BZ. In this section, we show that the Hall conductance of a filled band must be an integer, which is called the Chern number. The Berry curvature is the curl of the Berry gauge field $\mathbf{A}$. The BZ is a torus; hence, it has no boundary. An application of Stokes theorem would then give the Hall conductance as an integral of the Berry gauge field over the boundary of the BZ, but since the latter has no boundary, $\sigma_{xy} = 0$ if $\mathbf{A}(k)$ is well defined in the whole BZ. Nonzero values of the Hall conductance (Chern number) are consequences of the nontrivial structure of the Berry vector potentialin particular, the fact that it has singularities at points in the BZ. A nonzero Chern number (Hall conductance) reflects the fact that we cannot choose a global gauge that is continuous and single valued over the entire BZ. Hence, a nonzero Chern number represents an obstruction to the application of Stokes theorem over the whole BZ.

Under a U(1) gauge transformation, the wavefunction of the ath energy level transforms as $|a,k\rangle \to e^{if(k)}|a,k\rangle$, where we consider $f(k)$ a smooth function over the whole BZ. The corresponding transformation on the Berry potential $A'(k) = A(k) + \partial f(k)$. Because we performed a gauge transformation, the Hall conductance, which as an observable quantity is gauge invariant, can just as well be calculated in this new gauge. The availability of an arbitrary gauge transformation naively implies that the overall phase of the wavefunction can be chosen arbitrarily by making a suitable gauge transformation. For example, one phase choice is to choose the function $f(k)$ such that it makes the first component of the vector $|a,k\rangle_1$ real. If the first component of $|a,k\rangle$ is nonzero,we can always pick a phase to gauge-transform and make it real by choosing ei $f(k) = ||a,k\rangle_1|/|a,k\rangle_1$. Call this wavefunction ψ_1.

However, if we could always find a smooth gauge, then the Hall conductance would always

vanish, by the preceding argument involving Stokes theorem. Thus, it hence must be true that there are certain cases in which we cannot pick a smooth gauge for our wavefunction In our specific gauge-smoothing procedure, it must be that we cannot pick a phase to make the first component real. This happens when the first component of the Bloch part of $|a,k\rangle$ vanishes at some points in the Brillouin zone.We denote the positions in the BZ of the zeros of the first component of the Bloch wavefunction by k_s, with $s = 1,..,N$; around them,we subsequently define small regions $R_s = \{k \in T^2_{BZ} || k - k_s| < \varepsilon, |a,k_s\rangle_1 = 0\}$. Inside those regions, we cannot pick a smooth wavefunction because our gauge-smoothing procedure fails. Hence, we pick a different phase convention inside these regions: for example, we fix the phase by saying that $|a,k_s\rangle_2$ is real. Once we pick the phases, the state is completely well defined over the patches that contain the zeros of the first component. Call this wavefunction ψ_2, for which the gauge smoothing just means multiplication by $e^{ig(k)} = ||a,k\rangle_2|/|a,k\rangle_2$. Obviously, $\psi 2$ is smoothly defined within the small circles (it is not well defined everywhere outside the circles the second component vanishes outside the circles), whereas ψ_1 is smoothly defined outside the circles and has ambiguities related to the vanishing of its first component in the circles. At the boundary between the two regions, the two wavefunctions are related by a gauge transformation:

$$\psi_2(k) = e^{i(g(k)-f(k))}\psi_1(k) = e^{i\chi(k)}\psi_1(k) \tag{161}$$

Hence, the Berry potentials for the two wavefunctions are themselves related by a gauge transformation:

$$\mathbf{A}_2(k) = \psi_2(k)\partial_k\psi_2(k) = \psi_1(k)\partial_k\psi_1(k) + i\partial\chi(k) = \mathbf{A}_1(k) + i\partial\chi(k) \tag{162}$$

The Hall conductance is gauge invariant, but if we want to obtain it from integrating the curl of the Berry vector potential over the BZ, we must be able to have smoothly

differentiable wavefunctions.We have these by patches, so we write

$$\sigma_{xy} = \frac{e^2}{h}\frac{1}{2\pi i}[\int_{T_B^2 Z - R_s} \partial \times \mathbf{A}_1(k) + \int_{R_s} \partial \times \mathbf{A}_2(k)] \tag{163}$$

$$= \frac{e^2}{h}\frac{1}{2\pi i}[\int_{\partial(T_B^2 Z - R_s)} d\mathbf{k} \cdot \mathbf{A}_1(k) + \int_{\partial R_s} d\mathbf{k} \cdot \mathbf{A}_2(k)] \tag{164}$$

$$= \frac{e^2}{h}\frac{1}{2\pi i}[\int_{\partial(-R_s)} d\mathbf{k} \cdot \mathbf{A}_1(k) + \int_{\partial R_s} d\mathbf{k} \cdot \mathbf{A}_2(k)] \tag{165}$$

$$= \frac{e^2}{h}\frac{1}{2\pi i}[\int_{\partial(R_s)} d\mathbf{k} \cdot (\mathbf{A}_2(k) - \mathbf{A}_1(k))] \tag{166}$$

$$= \frac{e^2}{h}\frac{1}{2\pi i}[\int_{\partial(R_s)} d\mathbf{k} \cdot i\partial\chi(k) = \frac{e^2}{h}n \tag{167}$$

where $n = \frac{1}{2\pi}\oint_{\partial R_s} d\mathbf{k} \cdot \partial\chi(k)$. is the winding number of the gauge transformation on the boundary of the piecewise definition of the wavefunctions. As a simple example, if we take the boundary of the R_s a perfect circle $\partial R_s = k_s + \varepsilon e^{i\theta}$, with $\theta \in [0, 2\pi]$, we would have

$$n = \frac{1}{2\pi}\oint_{\partial R_s} d(\varepsilon e^{i\theta})\frac{\partial\chi(k)}{\partial\varepsilon e^{i\theta}} = \frac{1}{2\pi}\int_0^{2\pi} d\theta\partial_\theta\chi(k_s + \varepsilon e^{i\theta}) \tag{168}$$

$$= \frac{1}{2\pi}[\chi(k_s + \varepsilon e^{i2\pi - 0^-}) - \chi(k_s + \varepsilon)] \tag{169}$$

Notice that n has to be an integer because, aswe complete a full path around each circle.

Hence, upon a full revolution around the point k_s , we necessarily have $\chi(k_s + \varepsilon e^{i2\pi - 0^-}) - \chi(k_s + \varepsilon) = 2n\pi$. Although the phase of the wavefunction is gauge variant, the total vorticity (winding number) is a gauge-invariant quantity. The positions of the vorticities in the BZ can be changed, for example, by picking different components of the Bloch state to gauge smoothen; we can even separate vorticities and create vorticities from vacuum as long as vorticity conservation is maintainedbut the sum of all vorticities in the BZ is constant and equal to the Chern number. It is a topological invariant since it cannot change upon any smooth deformation of the contour.

Fiber bundle and the Chern number

Let us finally discuss the topological properties of Hall conductance. The argument goes as follows. The boundary-condition angles $\theta_{x,y}$, being phases, are defined modulo 2π. Each choice of a boundary condition amounts to a choice of a point $\theta_{x,y}$ on the torus T^2 of boundary conditions. For each point θ_x, θ_y we have a unique eigenstate $|n(\theta_x, \theta_y)\rangle$ of the

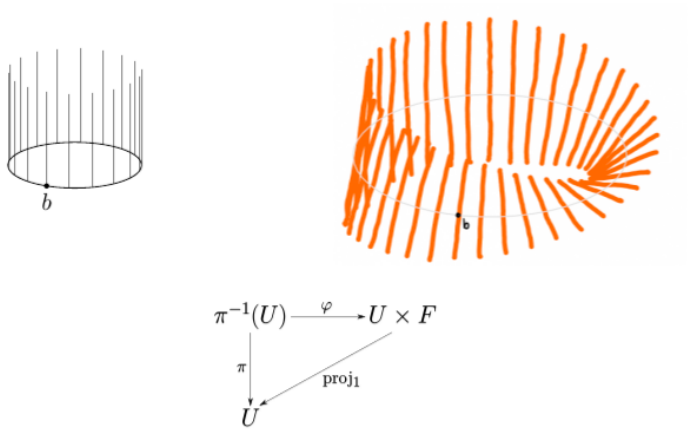

FIG. 13: A fiber bundle on S^1: 2 topological spaces.

full many-body Hamiltonian H. In mathematical jargon, we have a fiber bundle. The wave function has an amplitude and a phase that are smooth functions of θ_α. *The total phase of the wave function is not a physical observable, but changes of the phase are! (Quantum Mechanism!)* In particular, let us imagine that, at some initial time t_0, we have defined an initial boundary condition $\theta(t_0)$ with a phase for the state $arg|n(\theta(t_0))]$.

The fiber bundle associated with this problem can be defined in the following. With every point on T^2 we associate a state $|n(\theta)\rangle$. At every point $\theta \in T^2$ we have associated the ray or bundle of states related to $|n(\theta)\rangle$ by a gauge transformation. The torus T is partitioned into a union of sets T^I and T^{II}, . . . each containing at most one zero of $|n(\theta)\rangle$. The phase of $|n(\theta)\rangle$ is defined for each set, which results in a set of state vectors $|n^{I,II,\cdots}(\theta)\rangle$ whose phases are smoothly defined on T^I and T^{II}... These state vectors differ from each other just by gauge transformations that are smooth functions $f(\theta)$ on the overlap between two regions, say T^I and T^{II}. The transition function $f(\theta)$ is a smooth map from the closed curve $\gamma \in T^I \cap T^{II}$ to the group $U(1)$ of phases $e^{if(\theta)}$. These maps can be classified into homotopy classes, with each class defined by the winding number C_1. This map is known as the principal U(1) bundle over the torus T^2. The vector field $\mathbf{A}(\theta)$ defines a connection.

Lets define now the 1-form $dA = A_k d\theta_k$. A connection 1-form can be written as $\Omega = A + dA$. The transition functions act on fibers (i.e. state vectors) by multiplication. Once a connection A_k has been given, a curvature 2-form $F = dA$ can be defined, and it is known as the first Chern form. The integral of this 2-form is the first Chern number.

FIG. 14: The genus of sphere is 0, torus is 2 and double torus is 4. The related Euler characterics is 2, 0, -2, respectively.

In mathematics, and particularly topology, a fiber bundle is a space that is locally a product space, but globally may have a different topological structure. Specifically, the similarity between a space E and a product space is defined using a continuous surjective map $\pi : E \to B$. In the trivial case, E is just $B \times F$, and the map π is just the projection from the product space to the first factor. This is called a trivial bundle. Examples of non-trivial fiber bundles include the Mbius strip and Klein bottle.

We require that for every $x \in B$, there is an open neighborhood $U \in B$ of x (which will be called a trivializing neighborhood) such that there is a homeomorphism $\varphi : \pi^{-1}(U) \to U \times F$ (where $U \times F$ is the product space) in such a way that π agrees with the projection onto the first factor. That is, the following diagram should commute, where $proj_1 : U \times F \to U$ is the natural projection and $\varphi : \pi^{-1}(U) \to U \times F$ is a homeomorphism. The set of all $\{(U_i, \varphi_i)\}$ is called a local trivialization of the bundle.

Perhaps the simplest example of a nontrivial bundleE is the Mbius strip. It has the circle that runs lengthwise along the center of the strip as a base B and a line segment for the fiber F, so the Mbius strip is a bundle of the line segment over the circle.

A homeomorphism φ exists that maps the preimage of U (the trivializing neighborhood) to a slice of a cylinder: curved, but not twisted. This pair locally trivializes the strip. The corresponding trivial bundle $B \times F$ would be a cylinder, but the Mbius strip has an overall "twist". This twist is visible only globally; locally the Mbius strip and the cylinder are identical (making a single vertical cut in either gives the same space).

GaussBonnet theorem, Topology and Geometry

From the mathematical point of the view, the Berry curvature and the Gaussian curvature (geometry), are described by the same mathematical structure: fiber bundles. And shows that they are all quantized due to topological reasons, which is known as topological quantization.

$$\frac{1}{2\pi}\int KdS = \chi_M \tag{170}$$

$$\frac{1}{2\pi}\int BdS = C \tag{171}$$

The first one known as the Euler characteristic, which measuresthe topological nature of the manifold M, and the second one known as the Chern number, which measures the quantized Hall conductivity.

The idea of topology originates from geometry in the descriptions of manifolds in a 3-dimensional space. Later, it is generalized other dimensions and generic abstract space (including the Hilbert space in quantum physics). In geometry, if an manifold A can be adiabatically deformed into B, we said that they have the same topology. Otherwise, we say that they are topologically different. Examples: the surface of a sphere and the surface of a cube are topologically equivalent. But the surface of a sphere and the surface of donut (torus) are topologically different.

To distinguish different manifolds, mathematicians developed an object, which is called an index (topological index). It is a number. For objects with the same topology, the index takes the same value. Otherwise, the value will be different. For 2D closed manifold, the index is the Euler characteristic (Gauss-Bennet theorem):

$$\frac{1}{2\pi}\int KdS = \chi_M = 2(1-g) \tag{172}$$

To define the curvature for a curve, we use a circle to fit the curve around one point on the curve. The inverse radius =1/R gives us the curvature. For a manifold, one can draw lots of curves at one point. And one can get the curvature for all of these curves. Among all these curvatures, the largest and smallest one are known as principle curvatures $_1$ and $_2$. The Gaussian curvature is the product of they two. For a sphere, $_1 = _2 = 1/R$, so $K = 1/R^2$. For any 2D closed manifold, the integral of K over the manifold is always an even integer and

it value is invariant no matter how one deform the manifold (adiabatically). χ_M only cares about the topology of the manifold M. If we deform any manifold adiabatically (without changing the topology), χ_M will remain the same. If one change the topology, χ_M takes a different value. For example, Sphere : $chi_M = 2$, Torus : $\chi_M = 0$, double torus : $\chi_M = -2$.

Importantly, χ_M is directly related to the genus g of the manifold. The genus measure the number of "handles" on a object. A sphere has no Handel, so g = 0. For torus g=1. For double torus g=2. A coffee mug has one Handel. So a coffee mug is a torus from the topological point of the view, i.e. a coffee mug=a donut.

χ_M, by definition it describe the topology of polyhedrons. For Euler characteristic,

$$\chi_M = V - E + F \tag{173}$$

where V, E, and F are the numbers of vertices (corners), edges and faces respectively. For polyhedron cubic, they are topologically equivalent to a sphere, which has $\chi_M = 2$. So, these polyhedrons has $V - E + F = 2$.

Does topology have application in physics? Before 1960s, people thought the answer is no.

Short summary of berry phase

- Berry phase is a geometric phase, generating by the dynamical adiabatic process.
- Berry phase is described by the integral of Berry connection (gauge dependent) or Berry curvature (gauge independent). Berry connection describes the phase change of the wave function.
- For a closed surface enclosing a magnetic monopole, no matter what gauge one uses, the Berry connection (vector potential) must have some singularities.
- Intrinsically, wave function needs to be multi-component, with the complex phase. The singularity usually comes from the zeros of wave function.
- The total phase of the wave function is not a physical observable, but changes of the phase are!

- The quantization of Hall conductance is guaranteed by its connection with winding number or Chern number. That is, integral of Berry curvature on a closed manifold is first Chern number.

- From the mathematical point of the view, the Berry curvature and the Gaussian curvature (geometry), are described by the same mathematical structure: fiber bundles. Thus, people link the Hall conductance with a topological reason.

Sum rule of Berry curvature

Let us recall the Berry curvature $\mathbf{B}^n = \nabla \times \mathbf{A}^n = \partial_\mu A^n_\nu - \partial_\nu A^n_\mu$. If we sum up all energy levels, the sum of the Berry curvatures are zero.

$$\mathbf{B}^{(n)} = -Im \sum_{n' \neq n} \langle \nabla n | n' \rangle \times \langle n' | \nabla n \rangle = -Im \sum_{n' \neq n} \frac{\langle n | \nabla H | n' \rangle \times \langle n' | \nabla H | n \rangle}{(E_n - E_{n'})^2} \tag{174}$$

Next we calculate

$$\begin{aligned} \sum_n \mathbf{B}^{(n)} &= -Im \sum_{n',n,n' \neq n} \frac{\langle n | \nabla H | n' \rangle \times \langle n' | \nabla H | n \rangle}{(E_n - E_{n'})^2} \\ &= -Im \sum_{n',n,n' \neq n} \frac{\langle n' | \nabla H | n \rangle \times \langle n | \nabla H | n' \rangle}{(E_{n'} - E_n)^2} = -\sum_n \mathbf{B}^{(n)} \end{aligned} \tag{175}$$

which is odd when switch $n \to n'$, so it must be zero.

Gauge dependent Berry connection

Let us recall the spin-1/2 in the external magnetic field. Previously we used the form of

$$|R, -\rangle = \begin{pmatrix} \sin\frac{\theta}{2} e^{i\phi} \\ -\cos\frac{\theta}{2} \end{pmatrix}, |R, +\rangle = \begin{pmatrix} \cos\frac{\theta}{2} e^{-i\phi} \\ \sin\frac{\theta}{2} \end{pmatrix} \tag{176}$$

Next we consider an alternative form by making a gauge transfermation:

$$|R', -\rangle = \begin{pmatrix} \sin\frac{\theta}{2} \\ -\cos\frac{\theta}{2} e^{-i\phi} \end{pmatrix}, |R', +\rangle = \begin{pmatrix} \cos\frac{\theta}{2} \\ \sin\frac{\theta}{2} e^{i\phi} \end{pmatrix} \tag{177}$$

We repeat the calculations and obtain

$$\begin{aligned} &\mathbf{A}' = A'_R \vec{e}_r + A'_\theta \vec{e}_\theta + A'_\phi \vec{e}_\phi \\ &A'_r = i\langle R', -|\partial_R|R', -\rangle = 0 \\ &A'_\theta = \frac{1}{R} i\langle R', -|\partial_\theta|R', -\rangle = 0, \\ &A'_\phi = \frac{1}{R\sin\theta} i\langle R', -|\partial_\phi|R', -\rangle = \frac{\cos^2\frac{\theta}{2}}{R\sin\theta} \\ &B_{\theta\phi} = \vec{e}_r \frac{1}{R\sin\theta}[\partial_\theta(\sin\theta A'_\phi) - \partial_\phi A'_\theta] + \vec{\theta}\frac{1}{R}[\frac{1}{\sin\theta}\partial_\phi A'_R - \partial_R(RA'_\phi)] + \vec{e}_\phi \frac{1}{R}[\partial_R(RA'_\theta) - \partial_\theta A'_R] = -\frac{1}{2}\frac{\mathbf{R}}{R^3} \end{aligned} \tag{178}$$

Thus we see, by choosing different gauge, the berry connection **A** is gauge dependent, but the Berry curvature is gauge in-dependent. This is consistent with the Fiber bundle theory.

Again, we can verify that, the winding number is equal to the gauge difference integral on the closed loop C:

$$\oint_C dk(\mathbf{A}' - \mathbf{A}) = \int_0^{2\pi} d\phi R \sin\theta [A_\phi - A'_\phi] = \int_0^{2\pi} d\phi R \sin\theta \frac{\cos^2\frac{\theta}{2} + \sin^2\frac{\theta}{2}}{R\sin\theta} = 2\pi \qquad (179)$$

Remark. Multi-components of wave function. It is a Berry phase effect, so by definition it is from relative phase of different component.

* Electronic address: zhuwei@westlake.edu.cn

[1] Prange and Girvin, The Quantum Hall Effect. Springer.

[2] E. Fradkin, Field Theories of Condensed Matter Physics. Cambridge University Press, 2013.

[3] David Tong, The Quantum Hall Effect.