1

Band structure theory: a story about the fountain of youth.

A few remarks

- Very old theory. The fundimental problems was clarified in 20s-30s.
- Very simple. Basically free fermions.
- We had thought that we understood it.
- But it gives us a big surprise in the last 30 years, especially in the last 5 years.

1.1. Born– Oppenheimer approximation

Wavefunction (N_e electrons and N_i ions)

$$\psi_e(r_1, r_2, r_3 \ldots r_{N_e}) = \psi_e(\{r_i\}) \tag{1.1}$$

$$\psi_i(R_1, R_2, R_3 \ldots R_{N_i}) = \psi_i(\{R_i\}) \tag{1.2}$$

$$H = H_e + H_i + H_{ee} + H_{ie} + H_{ii} \tag{1.3}$$

$$H_e = -\frac{1}{2\,m_e}\left(\partial_{r_1}{}^2 + \partial_{r_2}{}^2 + \ldots \partial_{r_N}{}^2\right) \tag{1.4}$$

$$H_i = -\frac{1}{2\,M_i}\left(\partial_{R_1}{}^2 + \partial_{R_2}{}^2 + \ldots \partial_{R_N}{}^2\right) \tag{1.5}$$

$$H_{ee} = \sum_{i<j} \frac{e^2}{|r_i - r_j|} \tag{1.6}$$

$$H_{ii} = \sum_{i<j} \frac{n^2\,e^2}{|R_i - R_j|} \tag{1.7}$$

$$H_{ie} = \sum_{i,j} \frac{n\,e^2}{|R_i - r_j|} \tag{1.8}$$

Ion mass are much larger than electrons.

So Ions moves much slower (if $K.E._{\cdot i} = K.E._{\cdot e}$)

For electrons, ions are "stationary", so we can treat ions as a static background.

$$H = H_e + V_i + H_{ee} \tag{1.9}$$

Ignore H_{ee} for now,

$$H = H_e + V_i \tag{1.10}$$

Without interactions between electrons. The question here is: How does one electron move in the presence of lattice background?

A few remarks

- Born– Oppenheimer approximation is an assumption. So does the assumption that one can ignore electron-electron interactions. There is no solid theortical proof it must be valid. It is just a simple point to start from, but whether it is correct or not need to be examined by experirments.
- Philosophy: Start from something simple (not necessarily well-justified theortically) and see if it can produce the correct results, wich agrees with experiments. If yes, we say that the simple assumption captures the fundimental physics here. If not, we need more sophisciated assumptions to understand this system.

1.2. Why free fermion?

1. Simple: a good starting point

2. In many metal or insulators, free electron theory works very well

- Free fermion theory can explain lots of phenomena
- On the quantitative level, it gives more accurate results than a naive interacting theory (the mean-field approximation)
- Landau fermion liquid theory. (Will be discussed latter)

1.3. Free fermions in vacuum

$$-\frac{1}{2\,m}\,\nabla^2\,\psi\left(\vec{r}\right) = \epsilon\,\psi\left(\vec{r}\right) \tag{1.11}$$

(Continuous) translational symmetry: the system (H) is invariant when we change $\vec{r}$ to $\vec{r} + \vec{a}$. Here $\vec{a}$ is an arbitrary vector.

A continuous symmetry implies a conservation law. Here, it is the momentum conservation $\left[\vec{p},\,H\right] = 0$. So we can find common eigenstates for $\vec{p}$ and H.

Eigenstates of $\vec{p}$: plane waves

$$\psi_k = \exp\left(i\,\vec{k}\cdot\vec{r}\right) \tag{1.12}$$

so,

$$-\frac{1}{2\,m}\,\nabla^2\,\psi_k = \frac{k^2}{2\,m}\,\psi_k \tag{1.13}$$

Eigen-energy:

$$\epsilon_k = \frac{k^2}{2\,m} \tag{1.14}$$

1.4. Free fermions in a crystal

Crystal: atoms form a periodic structure (a lattice).

Fermions in a crystal: they will see a periodic potential.

$$\left[-\frac{1}{2\,m}\,\nabla^2 + V\left(\vec{r}\right)\right]\psi\left(\vec{r}\right) = \epsilon\,\psi\left(\vec{r}\right) \tag{1.15}$$

with

$$V\left(\vec{r}\right) = V\left(\vec{r} + \vec{a_i}\right) \tag{1.16}$$

Symmetry: discrete translational symmetry: the system (H) is invariant when we move the system by one lattice spacing $\vec{a_i}$

- Q: What is the eigenstate of this Hamiltonian?
- A: Bloch waves

(Here, I consider a 1D system for simplicity. All the conclusions discussed here can be generalized to other dimensions easily.)

$$\psi_{n,k}(r) = u_{n,k}(r) \exp(i\, k\, r) \tag{1.17}$$

Here $u_{n,k}$ is a periodic function of r with the same periodicity as the lattice

$$u_{n,k}(r + a) = u_{n,k}(r) \tag{1.18}$$

n is an integer (the band index) and k is a real number (the momentum).

Remarks:

- although k is called a "momentum". It is NOT the momentum of the Bloch wave. $k \neq \langle \psi_{n,k} | -i\, \partial_r | \psi_{n,k} \rangle$
- $-\frac{\pi}{a} < k < \frac{\pi}{a}$, where a is the lattice spacing. The region between $-\frac{\pi}{a}$ and $\frac{\pi}{a}$ is called the Brillouin zone
  - Q: Why k is confined to the BZ between $-\frac{\pi}{a}$ and $\frac{\pi}{a}$? What if $k > \frac{\pi}{a}$
  - A: $k + 2\pi/a$ is the same as k, if we redefine u

$$\psi_{n,k+2\pi/a}(r) = u_{n,k+2\pi/a}(r) \exp[i(k + 2\pi/a)\, r] = u_{n,k+2\pi/a}(r) \exp\!\left(i\,\frac{2\pi}{a}\, r\right) \exp(i\, k\, r) \tag{1.19}$$

Notice that both $u_{n,k}(r)$ and $\exp(i\, 2\pi\, r/a)$ are periodic functions of r with periodicity a (a lattice spacing), we can combine them together and define their product as $\tilde{u}(r)$ a new periodic function with the same periodicity.

$$\psi_{n,k+2\pi/a}(r) = u_{n,k}(r) \exp(i\, k\, r) = \psi_{n,k}(r) \tag{1.20}$$

Bottom line: $k + 2\pi/a$ is the same as k. So we just need the region between $-\frac{\pi}{a}$ and $\frac{\pi}{a}$.

How about the eigennergy ϵ?

Define eigen-energy of the state $\psi_{n,k}(r)$: $\epsilon_n(k)$

For each n, $\epsilon_n(k)$ as a function of k is called an energy band.

- $\epsilon_n(k)$ is a smooth function of k
  - Why?
  - A: $\partial_k \psi_{n,k}(r)$ has no singularity.
    - This was a major question of solid state physics. We understood it in the 20s-30s. Now we took it as granted, so this discussion almost totally disappeared from modern literature and textbooks (until recently!)
    - The study about $\partial_k \psi_{n,k}(r)$ created the foundation for topological insulators. We will see how important $\partial_k \psi_{n,k}$ is for TIs, later when we discuss about TIs.
    - It is also useful for graphene
- $\epsilon_n(k)$ usually don't cross with each other: (que to avoided level crossing. For more information, please refer to your favorate quantum mechanics textbook on "degenerate perturbation").
  - Different k may have the same energy $\epsilon_n(k) = \epsilon_m(k')$. These degeneracy is protected by the discrete translational symmetry. $H_{\text{hyber}} = |\psi_{n,k}\rangle\langle\psi_{m,k'}|$ is not allowed, because it breaks the translational symmetry.
  - But for same k, the energies of different bands cannot be the same.
- But! there are exceptions (sometimes exceptions usually lead to Noble Prize): Very famous example, graphene

1.5. Band structure theory and transport: metal, insulator, and semi-metal

Fermi statistics: At low T, fermions occupies all the states below the chemical potential. Two possibilities:

1. chemical potential cross with some of the bands

2. chemical potential doesn't cross with any band (μ is located inside the energy gap)

For case 1:

- Some bands are fully filled: valence bands
- Some bands are empty: empty bands

- Some bands are partially filled: conducting bands----- these bands give us conductivity

Apply a electric field (increase μ on one side and reduce the chemical potential on the other side), what will happen?

The side with a larger μ has a more electrons (higher density). The side with smaller μ has less electrons (lower density).

Density difference gives current! j=σ E

It is a metal.

- Fermi surface: Fermi surface controls all low energy physics.

For case 2:

- Some bands are fully filled: valence bands
- Some bands are empty: empty bands
- no partially filled: no conducting bands----- no conductivity

Apply a electric field (increase μ on one side and reduce the chemical potential on the other side), what will happen?

Number of electrons doesn't change (unless electric potential μ is tooooo large) no matter how one increases or decreases μ.

No density difference. No currents. j=0

It is an insulator.

- No Fermi surface

Case 3: semi-metal: the band structure has a band crossing point and the Fermi energy is located at the band crossing point (graphene)

- Same as an insulator: no conducting bands
- Same as a metal: it is a conductor

Bottom line: a metal with no Fermi surface

No Fermi surface is a big deal. Because the size of the Fermi surface (k_F) appears almost everywhere (heat capacity, magnetic susceptibility, density of the state ...). Not k_F is zero! One cannot just apply old knowledge on a metal with finite k_F to graphene. Everything needs to be recalculated. Some of them question are still under debating today in the study of graphene.

Case 4: Topological insulator: insulator inside, but metal outside.

2

Topological insulator part I: Phenomena

(Part II and Part III discusses how to understand a topological insluator based band-structure theory and gauge theory)

(Part IV discusses more complicated T-reversal invariance topological insulators)

One phenomena: bulk insulator: surface/edge is metallic.

Example: integer quantum Hall effect

2.1. Quantum Hall effect

Hall effect:
- Electrons moving in a 2D plane (2D electron gas or 2DEG): Apply an electric field E, we get some current j. Typically, $E \parallel J$. Let's assume they are in the x direction E_x and j_x. Resistivity $\rho_{xx} = E_x / j_x$
- Apply a B field perpendicular to the 2D plane. Lorentz force leads to charge accumulation for the top and bottom edge. This gives a E field perpendicular to the current (E_y). Hall resistivity: $\rho_{xy} = E_y / j_x$
- Rotational symmetry: $\rho_{xx} = \rho_{yy}$ and $\rho_{xy} = \rho_{yx}$
- Classical mechanics: $\rho_{xy} = B$

$$e\, E_y = e\, v\, B \tag{2.1}$$

$$j_x = e\, v\, n \tag{2.2}$$

$$\rho_{xy} = E_y / j_x = \frac{v\, B}{e\, v\, n} = \frac{B}{e\, n} \tag{2.3}$$

Remark: very useful experimental technique. It determines the carrier density.

Integer Quantum Hall effect

ρ_{xy}: plateaus. $\sigma_{xy} = \frac{1}{\rho_{xy}}$ is quantized: $\nu\, e^2 / h$ (ν is an integer, known as the filling factor or filling fraction). In the same time, $\rho_{xx} = 0$ at these plateaus.

Between two plateaus, not universal (sample dependent).

σ_{xy} quantization. Very accurate. The second best way to measure the fine structure constant

The fine structure constant is one of the key fundamental constants. In CGS unit

$$\alpha = \frac{e^2}{\hbar\, c} \tag{2.4}$$

c is the speed of light, whose value is exactly known (no error). Quantum Hall effect gives e^2 / h, so we can get α.

The most accurate way to determine α comes from g-2

2

Topological insulator part I: Phenomena

(Part II and Part III discusses how to understand a topological insluator based band-structure theory and gauge theory)

(Part IV discusses more complicated T-reversal invariance topological insulators)

One phenomena: bulk insulator: surface/edge is metallic.

Example: integer quantum Hall effect

2.1. Quantum Hall effect

Hall effect:
- Electrons moving in a 2D plane (2D electron gas or 2DEG): Apply an electric field E, we get some current j. Typically, $E \; // \; J$. Let's assume they are in the x direction E_x and j_x. Resistivity $\rho_{xx} = E_x / j_x$
- Apply a B field perpendicular to the 2D plane. Lorentz force leads to charge accumulation for the top and bottom edge. This gives a E field perpendicular to the current (E_y). Hall resistivity: $\rho_{xy} = E_y / j_x$
- Rotational symmetry: $\rho_{xx} = \rho_{yy}$ and $\rho_{xy} = \rho_{yx}$
- Classical mechanics: $\rho_{xy} = B$

$$e \, E_y = e \, v \, B \tag{2.1}$$

$$j_x = e \, v \, n \tag{2.2}$$

$$\rho_{xy} = E_y / j_x = \frac{v \, B}{e \, v \, n} = \frac{B}{e \, n} \tag{2.3}$$

Remark: very useful experimental technique. It determines the carrier density.

Integer Quantum Hall effect

ρ_{xy}: plateaus. $\sigma_{xy} = \frac{1}{\rho_{xy}}$ is quantized: $v \, e^2 / h$ (v is an integer, known as the filling factor or filling fraction). In the same time, $\rho_{xx} = 0$ at these plateaus.

Between two plateaus, not universal (sample dependent).

σ_{xy} quantization. Very accurate. The second best way to measure the fine structure constant

The fine structure constant is one of the key fundamental constants. In CGS unit

$$\alpha = \frac{e^2}{\hbar \, c} \tag{2.4}$$

c is the speed of light, whose value is exactly known (no error). Quantum Hall effect gives e^2 / h, so we can get α.

The most accurate way to determine α comes from g-2

Magnetic moment of an electron:

$$\mu = g\,\mu_B\,S/\hbar \tag{2.5}$$

where μ_B is a Bohr magneton and S is the spin of an electron.

Dirac equation tells us that $g = 2$, but in reality, $g = 2.0023193043622$

This difference is known as "anomalous magnetic moment". This anomaly is due to the interaction between the gauge fields and the particles (Dirac equation only describes the motion of a fermoin without gauge fields), which renormalized the value of g slightly away from 2.

$$a = \frac{g-2}{2} \tag{2.6}$$

This renormalization of g can be computed theoretically in Quantum electrodynamics (QED) using loop expansions. According to QED, a is directly related to the fine structure constant.

$$a = \frac{\alpha}{2\,\pi} + \sharp 2\,\alpha^2 + \sharp 3\,\alpha^3 + \sharp 4\,\alpha^4 + ... \tag{2.7}$$

Currently, we know the first 4 coefficients. However, because α is very small $\sim \frac{1}{127}$, the contributions from the higher order terms are very small (in the order of 10^{-10}). If we ignore the higher order terms, we can get α from the value of a.

Remarks:

- For measuring α, g-2 gives a more accurate result than IQH. the error bar from g-2 is 0.37×10^{-9}. Using IQH, it is 24×10^{-9}
- On the other hand, in g-2, one need to relies on QED to get α. But IQH offers an measurement of α independent of QED.
- The fact that these two measurements agree with each other offers a direct check to the theory of QED.
- Finally, we condensed matter physicists have something that almost as accurate as particle physics.

Difference between theories and experiments in condensed matter physics, typically: a few percent to a few hundred percent (can be larger). Much larger than particle physics

- Strongly correlated (cannot trust perturbation theory, in direct contrast to QED where $\alpha \ll 1$):
- QED: the strength of interactions (energy scale of interaction/energy scale of single-particle kinetic energy)

$$\alpha = \frac{e^2}{\hbar\,c} \sim 1 \Big/ 127 \tag{2.8}$$

- CMP: the strength of interactions (energy scale of interaction/energy scale of single-particle kinetic energy)

$$\alpha_{CMP} = \frac{e^2}{\hbar\,v_F} \tag{2.9}$$

- Fermi velocity v_F is typically 1/100-1/1000 of c

$$\alpha_{CMP} = \frac{e^2}{\hbar\,v_F} \sim 1 - 10 \tag{2.10}$$

- Perturbation theory cannot be used. one must need to compute all powers to get a reasonable answer, which is impossible
- Dirty background (impurities): extremely important for transport. σ is typically sample dependent (not universal). Depends on the density of impieties.

Why IQH is so accurate? (interactions and impurities do not matter). Please keep these questions and we will come back to them later, after we learn more about TIs.

Now, let's look at ρ_{xx}, which is 0 in a quantum Hall state.

- Q: Zero resistivity ρ_{xx}. A superconductor? A pefect metal?
- A: No. It is not a perfect conductor, not a superconductor. It is not even a conductor. It is an insulator.
- A clue: Conductivity is zero

Two formulars we are very familiar with:

$$j = \sigma\,E \tag{2.11}$$

$$E = \rho\,j \tag{2.12}$$

So:

$$\rho = 1/\sigma \tag{2.13}$$

This naive formula requires a very important assumption: j and E are in the same direction. If we apply E in the x direction, the current must also in the x direction and there cannot be any j_y, and vice versa.

This assumption is true sometimes, but not always

More generic formula:

$$\begin{pmatrix} j_x \\ j_y \\ j_z \end{pmatrix} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & \sigma_{zz} \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} \tag{2.14}$$

If all off-diagonal terms are zero ($\sigma_{xy} = \sigma_{yz} = \sigma_{zx} = 0$) and all the diagonal terms are the same we got:

$$j = \sigma E \tag{2.15}$$

$$\text{So } \rho = \frac{1}{\sigma} \tag{2.16}$$

Key: conductivity is a tensor (a $d \times d$ matrix). Not a scalar.

For 2D

$$\begin{pmatrix} j_x \\ j_y \end{pmatrix} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix} \begin{pmatrix} E_x \\ E_y \end{pmatrix} \tag{2.17}$$

- Q: How about resistivity?
- A: Of course, it is also a matrix

$$\begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} \rho_{xx} & \rho_{xy} \\ \rho_{yx} & \rho_{yy} \end{pmatrix} \begin{pmatrix} j_x \\ j_y \end{pmatrix} \tag{2.18}$$

$$\begin{pmatrix} E_x \\ E_y \end{pmatrix} = \begin{pmatrix} \rho_{xx} & \rho_{xy} \\ \rho_{yx} & \rho_{yy} \end{pmatrix} \begin{pmatrix} j_x \\ j_y \end{pmatrix} = \begin{pmatrix} \rho_{xx} & \rho_{xy} \\ \rho_{yx} & \rho_{yy} \end{pmatrix} \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix} \begin{pmatrix} E_x \\ E_y \end{pmatrix} \tag{2.19}$$

So

$$\begin{pmatrix} \rho_{xx} & \rho_{xy} \\ \rho_{yx} & \rho_{yy} \end{pmatrix} \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix} \begin{pmatrix} \rho_{xx} & \rho_{xy} \\ \rho_{yx} & \rho_{yy} \end{pmatrix} \tag{2.20}$$

The resitivity tensor is the (matrix) inverse of the conductivity tensor, and vice versa.

$$\begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix} = \begin{pmatrix} \rho_{xx} & \rho_{xy} \\ \rho_{yx} & \rho_{yy} \end{pmatrix}^{-1} = \frac{1}{\rho_{xx}\rho_{yy} - \rho_{xy}\rho_{yx}} \begin{pmatrix} \rho_{yy} & -\rho_{xy} \\ -\rho_{yx} & \rho_{xx} \end{pmatrix} \tag{2.21}$$

```
Inverse[ ρxx  ρxy ]
        [ ρyx  ρyy ]
```

$$\left\{ \left\{ \frac{\rho yy}{-\rho xy\,\rho yx + \rho xx\,\rho yy}, -\frac{\rho xy}{-\rho xy\,\rho yx + \rho xx\,\rho yy} \right\}, \left\{ -\frac{\rho yx}{-\rho xy\,\rho yx + \rho xx\,\rho yy}, \frac{\rho xx}{-\rho xy\,\rho yx + \rho xx\,\rho yy} \right\} \right\}$$

For a quantum Hall system, at each plateaus: $\rho_{xx} = \rho_{yy} = 0$

$$\begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix} = \frac{1}{-\rho_{xy}\rho_{yx}} \begin{pmatrix} 0 & -\rho_{xy} \\ -\rho_{yx} & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1/\rho_{yx} \\ 1/\rho_{xy} & 0 \end{pmatrix} \tag{2.22}$$

$$\sigma_{xx} = \sigma_{yy} = 0 \tag{2.23}$$

No conducivity. A insulator.

Between two IQH plateaus (ρ_{xx} has a peak),

$$\begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix} = \frac{1}{\rho_{xx}\,\rho_{yy} - \rho_{xy}\,\rho_{yx}} \begin{pmatrix} \rho_{yy} & -\rho_{xy} \\ -\rho_{yx} & \rho_{xx} \end{pmatrix} \tag{2.24}$$

Finite σ_{xx}. A metal. (It is a metal. σ_{xx} depends on impurity densities, non-universal, just like other metals).

2.2. Why a IQH state is an insulator: Landau levels

a charge neutral particle in 2D:

$$-i\,\partial_t\,\psi(x,\,y) = \left[\frac{1}{2\,m}\,(-i\,\hbar\,\partial_x)^2 + \frac{1}{2\,m}\,\left(-i\,\hbar\,\partial_y\right)^2\right]\psi(x,\,y) \tag{2.25}$$

$$H = \frac{1}{2\,m}\,(-i\,\hbar\,\partial_x)^2 + \frac{1}{2\,m}\,\left(-i\,\hbar\,\partial_y\right)^2 \tag{2.26}$$

a particle with charge e in 2D. Just change the momentum operator $\vec{p}$ into $\vec{p} + e\,\vec{A}\big/c$, where $\vec{A}$ is the vector potential, and change $-i\,\partial_t$ into $-i\,\partial_t - e\,\Phi/c$, where Φ is the Electric potential. (minimal coupling).

$$(-i\,\partial_t - e\,\Phi/c)\,\psi(x,\,y) = \left[\frac{1}{2\,m}\,\left(-i\,\hbar\,\partial_x - \frac{e}{c}\,A_x\right)^2 + \frac{1}{2\,m}\,\left(-i\,\hbar\,\partial_y - \frac{e}{c}\,A_y\right)^2\right]\psi(x,\,y) \tag{2.27}$$

$$-i\,\partial_t\,\psi(x,\,y) = \left\{\left[\frac{1}{2\,m}\,\left(-i\,\hbar\,\partial_x - \frac{e}{c}\,A_x\right)^2 + \frac{1}{2\,m}\,\left(-i\,\hbar\,\partial_y - \frac{e}{c}\,A_y\right)^2\right] + e\,\Phi/c\right\}\psi(x,\,y) \tag{2.28}$$

$$H = \frac{1}{2\,m}\,\left(-i\,\hbar\,\partial_x - \frac{e}{c}\,A_x\right)^2 + \frac{1}{2\,m}\,\left(-i\,\hbar\,\partial_y - \frac{e}{c}\,A_y\right)^2 + e\,\Phi/c \tag{2.29}$$

For us, E=0, so we can set Φ=0.

The vector potential satisfies

$$\nabla \times A = \partial_x\,A_y - \partial_y\,A_x = B \tag{2.30}$$

A is NOT a physical observable. For a fixed B field, A is not unique. If A is the vector potential for B ($\nabla \times A = B$), $A' = A + \nabla\,\chi$ is also the vector potential for B ($\nabla \times A' = B$). We can choose A arbitrarily and they all give the same physical results. This is known as a gauge choice.

For us, B is a constant (a uniform magnetic field), so we can choose

$$A_x = 0 \text{ and } A_y = B\,x \text{ (Landau gauge)} \tag{2.31}$$

$$A_x = -\frac{B\,y}{2} \text{ and } A_y = \frac{B\,x}{2} \text{ (symmetric gauge)} \tag{2.32}$$

Or any other gauge.

If we choose the Landau gauge,

$$H = \left[\frac{\hbar^2}{2\,m}\,(-i\,\partial_x)^2 + \frac{1}{2\,m}\,\left(-i\,\hbar\,\partial_y - \frac{e}{c}\,B\,x\right)^2\right] \tag{2.33}$$

- Q: Eigenstates of H?

H is invariant under translations along the y axis $y \to y' = y + \text{constant}$

So, $[p_y,\,H] = 0$

We can find common eigenstates for p_y and H. So the eigen states for H must take this form

$$\psi = f(x)\,\exp\left(-i\,k_y\,y\right) \tag{2.34}$$

$$H\,\psi = \epsilon\,\psi \tag{2.35}$$

$$-\frac{\hbar^2}{2\,m}\,f''(x) + \frac{1}{2\,m}\left(\hbar\,k_y - \frac{e}{c}\,B\,x\right)^2 f(x) = \epsilon\,f(x) \tag{2.36}$$

$$-\frac{\hbar^2}{2\,m}\,f''(x) + \frac{e^2\,B^2}{2\,m\,c^2}\left(x - \frac{c\,\hbar}{e\,B}\,k_y\right)^2 f(x) = \epsilon\,f(x) \tag{2.37}$$

A harmonic oscillator:

$$-\frac{\hbar^2}{2\,m}\,\frac{d}{d\,x^2}\,\phi(x) + \frac{k}{2}\,(x - x_0)^2\,\phi(x) = \epsilon\,\phi(x) \tag{2.38}$$

x_0 is the equilibrium position. k is the spring constant

Exactly the same.

$$x_0 = \frac{c\,\hbar}{e\,B}\,k_y \tag{2.39}$$

$$k = \frac{e^2\,B^2}{m\,c^2} \tag{2.40}$$

$$\psi_{n,k_y} = \phi_n(x - x_0)\,\exp\left(-i\,k_y\,y\right) \tag{2.41}$$

$$\epsilon_{n,k_y} = \left(n + \frac{1}{2}\right)\hbar\,\omega = \left(n + \frac{1}{2}\right)\hbar\,\sqrt{\frac{k}{m}} = \left(n + \frac{1}{2}\right)\frac{e\,B\,\hbar}{c\,m} \tag{2.42}$$

- ϵ as a function of k_y
  - Looks like energy bands (n--- band index and k_y--- the momentum)
  - All the states with same n (but different k_y) have the same energy (flat bands)
- μ in the gap: insulator (QH plateaus). μ cross with one of the bands (Landau levels): metal (between two plateaus)
- ϵ proportional to B
  - As B increases, the energy for each band increases continuously.
  - For a fixed μ, as B increases, μ in the gap $\to$ cross with Landau level n $\to$ in the gap $\to$ cross with n-1$\to$ in the gap ...

2.3. Edge states with Quantization of the Hall conductivity

Electrons at the edge can move along the edge in one direction (the direction is determined by the B field).

This edge states gives quantized Hall conductivity.

Top edge:

$$N = \int\frac{d\,p\,d\,r}{2\,\pi\,\hbar} = L\int\frac{d\,p}{2\,\pi\,\hbar} \tag{2.43}$$

W is the width of the sample

$$n = \frac{N}{L} = \int\frac{d\,p}{2\,\pi\,\hbar} \tag{2.44}$$

$$\delta n_{\text{top}} = \int_{k_F}^{k_{F\,\text{top}}}\frac{d\,p}{2\,\pi\,\hbar} = \frac{k_{F\,\text{top}} - k_F}{2\,\pi\,\hbar} = \frac{\mu_{\text{top}} - \mu}{2\,\pi\,\hbar\,v_{F\,\text{top}}} \tag{2.45}$$

net current for top edge:

$$I_{\text{top}} = e\,v_{F\,\text{top}}\,\delta n_{\text{top}} = e\,v_{F\,\text{top}}\,\frac{\mu_{\text{top}} - \mu}{2\,\pi\,\hbar\,v_{F\,\text{top}}} = e\,\frac{\mu_{\text{top}} - \mu}{2\,\pi\,\hbar} = e\,\frac{e\,V/2}{2\,\pi\,\hbar} = \frac{e^2}{2\,\pi\,\hbar}\,\frac{V}{2} \tag{2.46}$$

Similarly,

$$I_{\text{bottom}} = \frac{e^2}{2\pi\hbar} \frac{V}{2} \tag{2.47}$$

Notice that the current for both edges flows in the same directions! top edge: more particles moving to the right. So the direction of total current is right. Bottom edge: less particles moving to the left, so the net current is also to the right.

$$I = I_{\text{top}} + I_{\text{bottom}} = \frac{e^2}{2\pi\hbar} V = \frac{e^2}{h} V \tag{2.48}$$

Hall Conductance

$$\frac{1}{R_H} = \frac{I}{V} = \frac{e^2}{h} \tag{2.49}$$

Hall Conductivity:

$$\sigma = j/E = \frac{I}{L_y} \Big/ \frac{U}{L_y} = I/U = \frac{1}{R_H} = \frac{e^2}{h} \tag{2.50}$$

If one has n copies of edge states, each edge states will give

$$I_i = \frac{e^2}{2\pi\hbar} V \tag{2.51}$$

Total currents for all edge states

$$I = n\, I_i = n\, \frac{e^2}{h} V \tag{2.52}$$

Conductivity:

$$\sigma = \frac{I}{V} = n\, \frac{e^2}{h} \tag{2.53}$$

- Q: impurity scattering?
- A: No backward scatterings. (one-way current, no reflections)

This is one of the reason, why σ_{xy} is quantized so precisely. Impurities will not change it.

2.4. Why some insulators have chiral edge states?

A piece of plastic is an insulator. IQH is also an insulator. Why IQH has chiral edge state but a piece of plastic doesn't?

Puzzle to solve: Why insulators are different from each other? How many different types of insulators do we have in this world?

What tool should we use? Where should we start to look for an answer?

Hint 0: This is a quantum problem: σ_{xy} depends on $\hbar$.

- So we should look at quantum mechanics.

Hint 1: Transport is the motion of charge. In quantum mechanics, charge is closely related with one quantity.

Hint 1.5: They are conjugate variables to each other, like r and p.

- Phase. invariant under $\phi \to \phi + \delta\phi$ implies the conservation of charge.

Hint 2: Phase can often lead to quantization

- Wavefunction is single valued. If we go around a circle, the change of the phase can only be $2\pi n$ where n is an integer.
- Example: the quantization of magnetic flux.

Q: Phase of a wavefunction: important or not?

A: The relative phase (the phase difference between two states) is important (interference), but the absolute value of a phase is not important

If $|\psi\rangle$ is an eigenstate, $e^{i\phi}|\psi\rangle$ is the wavefunction for the same state. All the observables remains the same.

$\langle\psi|A|\psi\rangle = \langle\psi|e^{-i\phi}|A|e^{i\phi}|\psi\rangle$

A^+ answer:

The absolute value of a phase is not important, but the fact that phase is unimportant is VERY IMPORTANT.

3

Topological insulator part II: Berry Phase and Topological index

3.1. Last chapter

Topological insulator: an insulator in the bulk and a metal near the boundary (surface or edge)

Quantum Hall insulator: $\sigma_{xx} = 0$

- 2D metal (weak lattice or no lattice potential), $\epsilon = k^2 / 2\, m$. As long as $\mu > 0$, metal (assuming there is no impurities and no interactions)
- B field: Landau levels: $\epsilon_{n,k} = \left(n + \frac{1}{2}\right) \hbar\, \omega$ (very similar to energy bands). If μ in the gap, it is an insulator. If μ coincides with a Landau level, the system is a metal.
- These insulating states are topological insulators (bulk insulator with chiral edge states).

Chiral edge states gives quantized Hall conductivity $\sigma_{xy} = n\, e^2 / h$

- Q: Why is Hall conductivity nonzero? A: chiral edge states.
  - For a chiral edge state, the top and lower edges have opposite velocities (say electrons near the top edge goes to the right and the lower edge to the left).
  - If we apply +V/2 for top edge and -V/2 for bottom edge. We will have a higher electron density at the top edge and a lower density at the lower edge (because electrons like positive voltage). So we have more electrons moving to the right (top edge has more electrons) and less electrons moving to the left (low edge has less electrons). So there is a net current in the x direction.
  - Another way to see this: if we pass a current (from left to right), we need to put more charge near the top edge, and less charge near the bottom edge. Charge imbalance induces a E field in the y direction.
- Q: Why Hall conductivity is quantized? A: chiral edge states. One edge contributes to e^2 / h. If we have n edge states: $n\, e^2 / h$
- Q: Why are impurities unimportant? A: chiral edge states. Electrons cannot be reflected by an impurity.

3.2. Some insulators have chiral edge state but others don't. Why?

A piece of plastic is an insulator. a IQH state is also an insulator. Why does the IQH system have chiral edge states but a piece of plastic doesn't?

Q: Why insulators are different from each other? How many different types of insulators do we have in this world?

What tool should we use? Where should we start to look for an answer?

Hint 1: σ_{xy} depends on $\hbar$. So this phenomenon must be due to some quantum physics.

Hint 2: Transport is the motion of charge. In quantum mechanics, charge is closely related with one quantity. What is that?

- Hint 2.5: It is the conjugate variable of charge (like p is the conjugate variable of r)
- Phase: invariant under $\phi \rightarrow \phi + \delta\phi$ implies the conservation of charge.

Hint 3: Phase can often lead to quantization (e.g. quantization of $L_z = n\, \hbar$)

- Wavefunction is single valued. If we go around a circle, the change of the phase can only be $2\, \pi$ n where n is an integer.

Hint 4: Phase of a wavefunction: quantum physics. (This is good, because we know that the quantum Hall effect is a quantum phenomena).

So: we will focus on "phase".

- Q: Phase of a wavefunction: important or not?
- A: The relative phase (the phase difference between two states) is important (which leads to interference), but the absolute value of a phase is not important

If $|\psi\rangle$ is an eigenstate, $|\psi'\rangle = e^{i\phi}|\psi\rangle$ is the wavefunction for the same state. All the observables remains the same.

$$\langle\psi|A|\psi\rangle = \langle\psi|e^{-i\phi}|A|e^{i\phi}|\psi\rangle = \langle\psi'|A|\psi'\rangle \tag{3.1}$$

Relative phase is important

$$|\Psi\rangle = \frac{1}{\sqrt{2}}|\psi_1\rangle + \frac{1}{\sqrt{2}}|\psi_2\rangle \tag{3.2}$$

$$\langle\Psi|P|\Psi\rangle = \frac{\langle\psi_1|P|\psi_1\rangle + \langle\psi_2|P|\psi_2\rangle}{2} + \frac{\langle\psi_1|P|\psi_2\rangle + \langle\psi_2|P|\psi_1\rangle}{2} \tag{3.3}$$

The physical meaning of the first term: system has half of the probably on state $|\psi_1\rangle$ and half on $|\psi_2\rangle$. So the momentum of $|\Psi\rangle$ is the average of $|\psi_1\rangle$ and $|\psi_2\rangle$, just like in classical systems.

The physical meaning of the second term: interference between the two quantum state. (relies on the relative phase between state 1 and 2)

$$|\Psi'\rangle = \frac{1}{\sqrt{2}}|\psi_1\rangle + \frac{e^{i\phi}}{\sqrt{2}}|\psi_2\rangle \tag{3.4}$$

$$\langle\Psi'|P|\Psi'\rangle = \frac{\langle\psi_1|P|\psi_1\rangle + \langle\psi_2|P|\psi_2\rangle}{2} + \frac{e^{i\phi}\langle\psi_1|P|\psi_2\rangle + e^{-i\phi}\langle\psi_2|P|\psi_1\rangle}{2} = $$
$$\frac{\langle\psi_1|P|\psi_1\rangle + \langle\psi_2|P|\psi_2\rangle}{2} + \cos\phi\frac{\langle\psi_1|P|\psi_2\rangle + \langle\psi_2|P|\psi_1\rangle}{2} \tag{3.5}$$

Example: flux quantization

- an A^+ answer: The absolute value of a phase is not important, but the fact that phase is unimportant is VERY IMPORTANT.

3.3. Example: Central charge and quantum group

References:

- Weinberg, The quantum theory of fields, Vol. 1, Chapter 2.2 and 2.7.
- Polchinski, String theory, Vol. 1
- Di Francesco, Mathieu, Senechal, Conformal Field theory

Remarks on central charge:

- Not directly related to topological insulators
- Same physics principle: phase of a quantum state is unimportant, which is very important
- Useful when we discuss 1+1D quantum physics in later chapters
- It is always good to know the connections between different pieces of knowledge.

3.3.1. Commutation relations:

Commutation relations are crucial for quantum systems.

There are two types of commutation relations:

- Type 1: Obvious e.g. $[L_x, L_y] = i L_z$
- Type 2: Not so obvious: central charge

3.3.2. Example of the first type: $[L_x, L_y] = i L_z$

- Q: Why angular momentum operators don't commute with each other.

- A: Rotating a system along the x axis first then along the y axis is different from rotating a system along the y axis first then along the x axis

Angular momentum is relation with space rotation. (generators of rotations)

$$R_x(\theta) = \exp(-i \ \theta \ L_x) \tag{3.6}$$

$$R_y(\theta) = \exp(-i \ \theta \ L_y) \tag{3.7}$$

$$R_z(\theta) = \exp(-i \ \theta \ L_z) \tag{3.8}$$

$$L_x = i \ \partial_\theta R_x(\theta) \mid_{\theta \to 0} \tag{3.9}$$

$$L_y = i \ \partial_\theta R_y(\theta) \mid_{\theta \to 0} \tag{3.10}$$

$$L_z = i \ \partial_\theta R_z(\theta) \mid_{\theta \to 0} \tag{3.11}$$

All rotations form a "Group" (SO(3)), a set with product.

- Product: the product of two elements X and Y is another element of this set $Z = X Y$. (Closure)

- $X Y$ and $Y X$ don't need to be the same.

- For rotations, for any two rotations R_1 and R_2, their product $R_2 R_1$ is defined as we first do the rotation R_2 and then R_1, which is another rotation $R_3 = R_2 R_1$. For example $R_x(\theta_1) R_x(\theta_2) = R_x(\theta_1 + \theta_2)$

- Associativity: $A (B C) = (A B) C$

- Identity element I. For any element X, $I X = X I = X$.

- For rotations, I is rotation by zero degree. $I = R(0)$.

- Inverse element: For any element X, there is another element X^{-1} which satisfies $X X^{-1} = X^{-1} X = I$

- For rotations, the inverse of $R_x(\theta)$ is $R_x(-\theta)$, rotating along the same axis but we flip the sign of the angle.

For rotations, the order of the product matters.

$$R_x(\theta_1) R_y(\theta_2) \neq R_y(\theta_2) R_x(\theta_1) \tag{3.12}$$

So

$$\left[R_x(\theta_1), R_y(\theta_2) \right] \neq 0 \tag{3.13}$$

Rotations don't commute with each other. This tells us that their generators cannot commute with each other.

$$\left[L_x, L_y \right] \neq 0 \tag{3.14}$$

If $\left[L_x, L_y \right] = 0$, we will have $\left[\exp(-i \ \theta_1 L_x), \exp(-i \ \theta_2 L_y) \right] = 0$, which is obviously incorrect.

Using the commutation relation between rotations, we can find that:

$$\left[L_x, L_y \right] = i L_z \tag{3.15}$$

A geometry problem. No quantum physics invovled.

3.3.3. Example of the second type: Why is mass NOT an operator? Galilean transformation and space translation: commute or not?

Space translations: $T(a)$ changes $x \to x' = x + a$. The generator is the momentum operator p:

$$T(a) = \exp(-i \ p \ a) \tag{3.16}$$

$$p = i \ \partial_a T(a) \mid_{a \to 0} \tag{3.17}$$

Galilean transformation: $G(v)$ changes $x \to x' = x + v t$. The generator is called K.

$$G(v) = \exp(-i \ K \ v) \tag{3.18}$$

$$K = i \ \partial_v G(v) \mid_{v \to 0} \tag{3.19}$$

- Q: [K,p]=?
  - A: Classical physics and quantum physics have different answer.

Classical physics: $[T, G] = 0$ so [K,p]=0

$$G_v : x \to x' = x + v t \tag{3.20}$$

$$T_a : x \to x' = x + a \tag{3.21}$$

$$G_v \psi(x, t) = \psi(x + v t, t) \tag{3.22}$$

$$T_a G_v \psi(x, t) = T_a \psi(x + v t, t) = \psi(x + v t + a, t) \tag{3.23}$$

$$T_a \psi(x, t) = \psi(x + a, t) \tag{3.24}$$

$$G_v T_a \psi(x, t) = G_v \psi(x + a, t) = \psi(x + v t + a, t) \tag{3.25}$$

$$T_a G_v \psi(x, t) = G_v T_a \psi(x, t) \tag{3.26}$$

$$T_a G_v = G_v T_a \tag{3.27}$$

$$[T_a, G_v] = 0 \tag{3.28}$$

So $\tag{3.29}$

$$\partial_a \partial_v [T_a, G_v] = 0 \tag{3.30}$$

$$[\partial_a T_a, \partial_v G_v] = 0 \tag{3.31}$$

$$[i \partial_a T_a, i \partial_v G_v] = 0 \tag{3.32}$$

$$[K, p] = \lim_{a \to 0} \lim_{v \to 0} [i \partial_a T_a, i \partial_v G_v] = 0 \tag{3.33}$$

Quantum physics:

Galilean transformation change velocity from u to u+v. So it change momentum p to p+m v where m is the mass

$$G(v) p G^{-1}(v) = p + m v \tag{3.34}$$

$$\exp(-i K v) p \exp(i K v) = p + m v \tag{3.35}$$

At small v:

$$\exp(-i K v) = 1 - i v K + O(v^2) \tag{3.36}$$

$$\exp(i K v) = 1 + i v K + O(v^2) \tag{3.37}$$

$$(1 - i v K) p (1 + i v K) = p + m v \tag{3.38}$$

$$p - i v K p + i v p K = p + m v \tag{3.39}$$

$$-i K p + i p K = m \tag{3.40}$$

$$K p - p K = i m \tag{3.41}$$

$$[K, p] = i m \tag{3.42}$$

Q: Why do we have two different answers?

A: Phase is unimportant in QM, which is very important.

$$T_a G_v x = x + v t + a \tag{3.43}$$

$$G_v T_a x = x + a + v t \tag{3.44}$$

$$G_v T_a x = T_a G_v x \tag{3.45}$$

This equation tells us that $T_a G_v$ and $G_v T_a$ are the same transformation. For classical physics, it simply means $T_a G_v = G_v T_a$.

In quantum mechanics, $T_a G_v$ and $G_v T_a$ are the same transformation means that $T_a G_v \psi(x, t)$ and $G_v T_a \psi(x, t)$ must give the same quantum state. However, they don't need to have the same phase, because phase itself is unimportant. This means that

$$T_a G_v \psi(x, t) = e^{i \phi(v, a)} G_v T_a \psi(x, t) \tag{3.46}$$

$$T_a G_v = e^{i \phi(v, a)} G_v T_a \tag{3.47}$$

- Key: in quantum mechanics "the same" means equal up to a phase difference.

$$[T, G] = T_a G_v - G_v T_a = \left(e^{i \phi(v, a)} - 1\right) G_v T_a \neq 0, \text{ if the phase is nonzero } \phi(v, a) \neq 0. \tag{3.48}$$

The operations also form a group, but their products are different from classical physics. This is known as a quantum group (or projective representation).

For generators, this extra phase means that we need to add some extra piece to their commutation relations, which is known as central charge.

For K and p, this implies that [K,p]=$i C$, where C is known as a central charge. It is just the mass.

- The central charge needs to satisfy some self-consistency condition: X (Y Z)=(X Y) Z.
- For many groups, this condition implies C=0 (no extra piece. Same as classical physics.) Example: rotational groups SO(3), Lorentz group ...
- The central charge must commute with all other symmetry operators due to the self-consisstency condition, i.e. it is a number.

Because mass is a central charge, mass must commute with all other operators. In other words, mass is a number in nonrelastivitic QM.

Bottom line: phase is unimportant implies that we can have a central charge in the commutation relations. This is a quantum phenomena with no parallel example in classical physics.

3.3.4. A more important example of central charge: string theory and 1D quantum physics
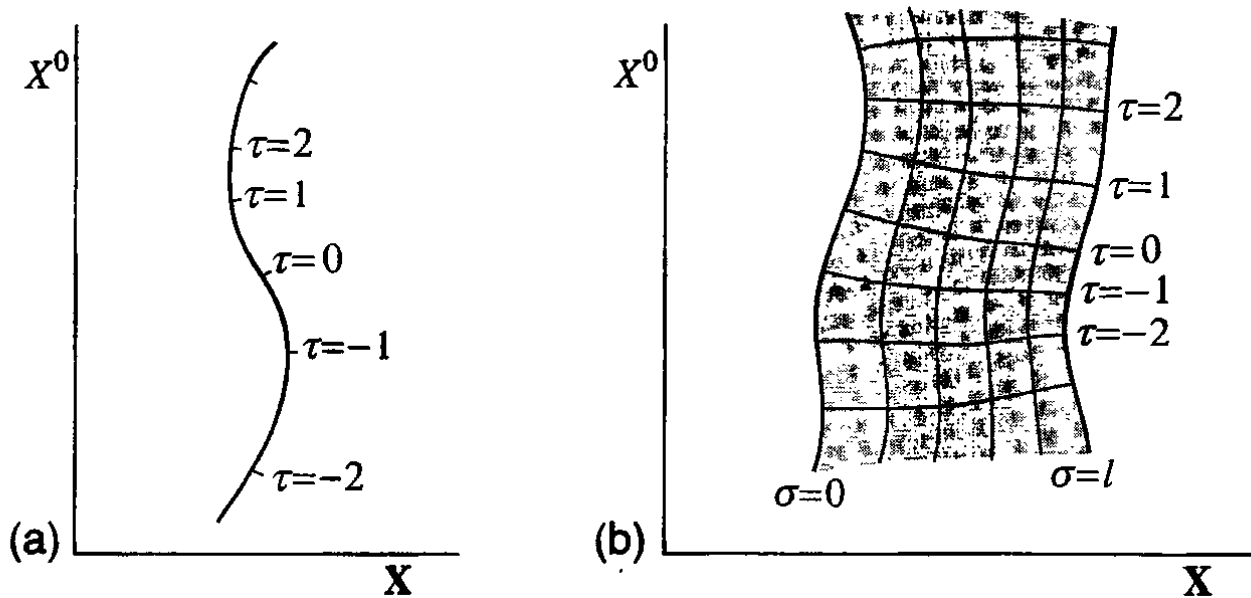

Figure from Polchinski, String theory, Vol. 1. (a) World-line of a particle and (b) world-sheet of a string.

For a particle moving in D-dimenstional space-time.

- Q: What is the trace of this particle?
- A: 1D curve: world-line.

$$X^\mu = X^\mu(\tau) \tag{3.49}$$

with μ=t, x, y, z and τ is a parameter along the world-line.

For a string moving in D dimensional space-time.

- Q: What is the trace of this string?

- A: 2D surface: world-sheet.

$$X^\mu = X^\mu(\sigma, \tau) \tag{3.50}$$

σ and τ are the coordinate on the 2D world sheet, which lives in a D dimensional space.

Quantum 1D system in condensed matter. Same description, same physics and same technique (conformal field theory). For example, density of electrons in a 1D Fermi gas:

$$\rho(x, t) \tag{3.51}$$

String theory and critical 1D quantum systems (including 1D free fermions or bosons) have a very special symmetry: 2D conformal symmetry.

Conformal symmetry:

for a 2D system, we can define the complex coordinate $z = x + i\, t$ (or $z = \sigma + i\, \tau$ for the string theory)

A conformal transformation: a mapping $z \to f(z)$ for any an analytic function $f(z)$ (functions which can be expand into a Taylor series, including most of the functions we usually use: $e^{\alpha z}$, $\sin k\, z$, z^n, $\log z$ ...).

- Conformal transformation preserves the angles between two vectors (general definition in any dimensions).

- In higher dimensions, d>2, there are only $\frac{(d+2)(d+1)}{2}$ of them.

- In d=2, there are infinite number of them. (infinite analytic functions)

- Translations are conformal transformations: $f(z) = z + x_0 + i\, t_0$

- Rotations are also conformal transformations: $f(z) = z\, e^{i\,\phi}$

- Dilation is also a conformal transformation: $f(z) = a\, z$

Conformal symmetry: systems invariant under all conformal transformations.

- Example: string theory or a 1D critical system

- Noether theorem: A symmetry implies one conservation law. Infinite conformal transformation, and the system is invariant under every one of them. This means that there are infinite numbers of conservations law.
  - This is why string theory is easy to handle.

Generators of conformal transformations: l_n with n an integer:

$$l_n = -z^{n+1}\, \partial_z \tag{3.52}$$

- $n = -1$ gives the generators of translations $l_{-1} = -\partial_z = \frac{-\partial_x + i\, \partial_y}{2}$. Real part is just $-i\, p_x/2$ and imaginary part gives $-i\, p_y/2$

- $n = 0$ gives the generators of rotations and dilations $l_0 = -z\, \partial_z = (x + i\, y)\, \frac{-\partial_x + i\, \partial_y}{2} = -(x\, \partial_x + y\, \partial_y) + i\, (x\, \partial_y - y\, \partial_x)$. Imaginary part is proportional to $L_z = r \times p = -i\, (x\, \partial_y - y\, \partial_x)$.

Classical physics

$$[l_m, l_n] = (m - n)\, l_{m+n} \tag{3.53}$$

Quantum physics: (Virasoro algebra)

$$[L_m, L_n] = (m - n)\, L_{m+n} + \frac{c}{12}\, (m^3 - m)\, \delta_{m,-n} \tag{3.54}$$

c is the central charge.

Similarities and differences between studying string theory and learning to play guitar?

- Similarities: both studies the motions of strings (the equations of motion are pretty much the same)

- Differences: the formal studies quantum strings (with central charge), while the latter handle classical strings (no central charge)

Central charge in the string theory: why D=26?

For X^μ, each of them has c=1 and there are D of them, where D is the dimension of the space. In addition, one need ghost fields for gauge symmetry, which also have central charge.

Total central charge C=D-26=0

D=26. This is why we should study bosonic string theory in 26 dimentions (supersymmetry has extra fermions, so the C=0 condition is different and one works in D=10).

Central charge in condensed matter systems: (one example)

In condensed matter physics, 1D Fermi gas and Bose gas has the same Central charge C=1. They are totally equivalent to each other. In 1D, one can turn bosons into fermions or fermions into bosons and vice versa, which is known as "bosonization".

For C<1, we know pretty much everything about the critical behavior (scaling laws) by knowing the value of C.

For C≥1, a little bit more complicated. Some additional symmetry could be helpful (affine Lie algebras)

3.4. Example II: Phase of Bloch waves

3.4.1. Phase of a Bloch wavefunction

A particle moving in a periodic potential:

$$\left[-\frac{1}{2m} \nabla^2 + V\left(\vec{r}\right) \right] \psi\left(\vec{r}\right) = \epsilon \, \psi\left(\vec{r}\right) \tag{3.55}$$

with

$$V\left(\vec{r}\right) = V\left(\vec{r} + \vec{a}_i\right) \tag{3.56}$$

Eigen-wavefunctions: Bloch waves

$$\psi_{n,k}(r) = u_{n,k}(r) \exp(i\,k\,r) \tag{3.57}$$

- Q: Phase of $u_{n,k}(r)$?
- A: Doesn't matter. Phase itself is unimportant.

Bloch waves are invariant under a global phase shift in the momentum space. (Same as other quantum states)

If $\psi_{n,k}(r) = u_{n,k}(r) \exp(i\,k\,r)$ is a Bloch wave, then $\psi'_{n,k}(r) = e^{i\phi} u_{n,k}(r) \exp(i\,k\,r)$ is the same eigen state

In addition, Bloch waves are invariant under a local phase shift in the momentum space. (new and important)

If $\psi_{n,k}(r) = u_{n,k}(r) \exp(i\,k\,r)$ is a Bloch wave, then $\psi'_{n,k}(r) = e^{i\phi(k)} u_{n,k}(r) \exp(i\,k\,r)$ is the same eigen state

$$\left[-\frac{1}{2m} \nabla^2 + V\left(\vec{r}\right) \right] \psi_{n,k}(r) = \epsilon \, \psi_{n,k}(r) \tag{3.58}$$

$$\left[-\frac{1}{2m} \nabla^2 + V\left(\vec{r}\right) \right] \psi'_{n,k}(r) = \left[-\frac{1}{2m} \nabla^2 + V\left(\vec{r}\right) \right] e^{i\phi(k)} \psi_{n,k}(r) =$$

$$e^{i\phi(k)} \left[-\frac{1}{2m} \nabla^2 + V\left(\vec{r}\right) \right] \psi_{n,k}(r) = e^{i\phi(k)} \epsilon \, \psi_{n,k}(r) = \epsilon \, e^{i\phi(k)} \psi_{n,k}(r) = \epsilon \, \psi'_{n,k}(r) \tag{3.59}$$

The system is invariant under a local phase shift in the momentum space.

Requirement to have such a local symmetry.

1 Momentum conservation (no impurities)

2. No electron-electron interactions

So, a particle with momentum k always has momentum k. Bloch waves with different momentum doesn't talk to each other. So their phase can change independently.

3.5. Local phase symmetry

3.5.1. How to describe this local phase degrees of freedom?

Q: Do we know something similar?

A: Yes, we know a similar local phase degrees of freedom. Not in the k-space, but in the real space, which is the gauge symmetry.

Central charge in condensed matter systems: (one example)

In condensed matter physics, 1D Fermi gas and Bose gas has the same Central charge C=1. They are totally equivalent to each other. In 1D, one can turn bosons into fermions or fermions into bosons and vice versa, which is known as "bosonization".

For C<1, we know pretty much everything about the critical behavior (scaling laws) by knowing the value of C.

For C≥1, a little bit more complicated. Some additional symmetry could be helpful (affine Lie algebras)

3.4. Example II: Phase of Bloch waves

3.4.1. Phase of a Bloch wavefunction

A particle moving in a periodic potential:

$$\left[-\frac{1}{2m}\nabla^2 + V\!\left(\vec{r}\right)\right]\psi\!\left(\vec{r}\right) = \epsilon\,\psi\!\left(\vec{r}\right) \tag{3.55}$$

with

$$V\!\left(\vec{r}\right) = V\!\left(\vec{r} + \vec{a}_i\right) \tag{3.56}$$

Eigen-wavefunctions: Bloch waves

$$\psi_{n,k}(r) = u_{n,k}(r)\exp(i\,k\,r) \tag{3.57}$$

- Q: Phase of $u_{n,k}(r)$?
- A: Doesn't matter. Phase itself is unimportant.

Bloch waves are invariant under a global phase shift in the momentum space. (Same as other quantum states)

If $\psi_{n,k}(r) = u_{n,k}(r)\exp(i\,k\,r)$ is a Bloch wave, then $\psi'_{n,k}(r) = e^{i\phi}\,u_{n,k}(r)\exp(i\,k\,r)$ is the same eigen state

In addition, Bloch waves are invariant under a local phase shift in the momentum space. (new and important)

If $\psi_{n,k}(r) = u_{n,k}(r)\exp(i\,k\,r)$ is a Bloch wave, then $\psi'_{n,k}(r) = e^{i\phi(k)}\,u_{n,k}(r)\exp(i\,k\,r)$ is the same eigen state

$$\left[-\frac{1}{2m}\nabla^2 + V\!\left(\vec{r}\right)\right]\psi_{n,k}(r) = \epsilon\,\psi_{n,k}(r) \tag{3.58}$$

$$\begin{aligned}
&\left[-\frac{1}{2m}\nabla^2 + V\!\left(\vec{r}\right)\right]\psi'_{n,k}(r) = \left[-\frac{1}{2m}\nabla^2 + V\!\left(\vec{r}\right)\right]e^{i\phi(k)}\,\psi_{n,k}(r) = \\
&e^{i\phi(k)}\left[-\frac{1}{2m}\nabla^2 + V\!\left(\vec{r}\right)\right]\psi_{n,k}(r) = e^{i\phi(k)}\,\epsilon\,\psi_{n,k}(r) = \epsilon\,e^{i\phi(k)}\,\psi_{n,k}(r) = \epsilon\,\psi'_{n,k}(r)
\end{aligned} \tag{3.59}$$

The system is invariant under a local phase shift in the momentum space.

Requirement to have such a local symmetry.

1 Momentum conservation (no impurities)

2. electron interactions

So, a particle with momentum k always has momentum k. Bloch waves with different momentum doesn't talk to each other. So their phase can change independently.

3.5. Local phase symmetry

3.5.1. How to describe this local phase degrees of freedom?

Q: Do we know something similar?

A: Yes, we know a similar local phase degrees of freedom. Not in the k-space, but in the real space, which is the gauge symmetry.

This section, I review some basic knowledge we learned in the quantum mechanics class.

Reference: Landau and Lifshitz, Quantum Mechanics (non-relativistic theory).

3.5.2. Global $U(1)$ phase symmetry in real space and particle conservation

Schrodinger equation is invariant under global phase shift in the real space.

$$i \frac{\partial}{\partial t} \psi(r, t) = -\frac{\nabla^2}{2m} \psi(r, t) + V(r) \psi(r, t) \tag{3.60}$$

Define: $\psi'(r) = e^{i\phi} \psi(r)$ where ϕ is a constant (We shift the phase of every point by the same amount, which is known as a global phase change)

$$i \frac{\partial}{\partial t} \psi'(r, t) = i \frac{\partial}{\partial t} \left[e^{i\phi} \psi(r, t) \right] =$$

$$e^{i\phi} \left[i \frac{\partial}{\partial t} \psi(r, t) \right] = e^{i\phi} \left[-\frac{\nabla^2}{2m} \psi(r, t) + V(r) \psi(r, t) \right] = -\frac{\nabla^2}{2m} e^{i\phi} \psi(r, t) + V(r) e^{i\phi} \psi(r, t) = = -\frac{\nabla^2}{2m} \psi'(r, t) + V(r) \psi'(r, t) \tag{3.61}$$

- Key: ϕ is a constant, so that $\frac{\partial}{\partial t} e^{i\phi} = e^{i\phi} \frac{\partial}{\partial t}$ and $\nabla e^{i\phi} = e^{i\phi} \nabla$

$\psi'(r)$ follows the same equation as $\psi(r)$. The system is invariance under a global phase shift, which is known as a global U(1) phase symmetry.

Symmetry implies conservation law. Here, this symmetry means particle conservation (total probability density of a wavefunction is conserved $\partial_t \int dr \langle \psi | \psi \rangle = 0$).

Demonstration (no a proof)

Translational symmetry $x \to x + a$ implies the conservation of $-i \partial_x$. So here, for $U(1)$ phase symmetry, we expect the conservation of $-i \partial_\phi$. Here I show that $\int dr \langle \psi | -i \partial_\phi | \psi \rangle = \int dr \langle \psi | \psi \rangle$. So the conservation of $-i \partial_\phi$ implies the conservation of total probability density.

For a wavefunction $\psi(r, t)$

$$-i \partial_\phi \psi(r, t) = -i \frac{\psi(r, t) e^{i\delta\phi} - \psi(r, t)}{\delta\phi} = -i \frac{e^{i\delta\phi} - 1}{\delta\phi} \psi(r, t) = -i \frac{(1 + i\delta\phi) - 1}{\delta\phi} \psi(r, t) = \psi(r, t) \tag{3.62}$$

$$\langle \psi | -i \partial_\phi | \psi \rangle = \int dr\, \psi^*(r, t) \left[-i \partial_\phi \psi(r, t) \right] = \int dr\, \psi^*(r, t) \psi(r, t) = \int dr\, (| \psi(r, t) |)^2 \tag{3.63}$$

So $\int dr\, (| \psi(r, t) |)^2$ is conserved.

Here, the phase change ϕ is a constant, instead of an arbitrary function. What will happen if ϕ is a function of r?

3.5.3. Local $U(1)$ phase symmetry in real space and gauge fields

Define: a local phase shift $\psi'(r, t) = e^{i\phi(r,t)} \psi(r, t)$

Q: Is Schrodinger equation invariant under local phase shift?

A: No.

Key: ϕ is a function of t and r, so that $\frac{\partial}{\partial t} e^{i\phi} \neq e^{i\phi} \frac{\partial}{\partial t}$ and $\nabla e^{i\phi} \neq e^{i\phi} \nabla$

$$i \frac{\partial}{\partial t} \psi'(r, t) = i \frac{\partial}{\partial t} \left[e^{i\phi(r,t)} \psi(r, t) \right] =$$

$$-\frac{\partial\phi(r, t)}{\partial t} e^{i\phi(r,t)} \psi(r, t) + i e^{i\phi(r,t)} \frac{\partial}{\partial t} \psi(r, t) = -\frac{\partial\phi(r, t)}{\partial t} e^{i\phi(r,t)} \psi(r, t) - e^{i\phi(r,t)} \frac{\nabla^2}{2m} \psi(r, t) + e^{i\phi(r,t)} V(r) \psi(r, t)$$

$$= -\frac{\partial\phi(r, t)}{\partial t} \psi'(r, t) - e^{i\phi(r,t)} \frac{\nabla^2}{2m} \left[e^{-i\phi(r,t)} \psi'(r, t) \right] + V(r) \psi'(r, t) = \tag{3.64}$$

$$-\frac{\partial\phi(r, t)}{\partial t} \psi'(r, t) - \frac{\nabla^2}{2m} \psi'(r, t) + V(r) \psi'(r, t) + \frac{i}{m} [\nabla \phi(r, t) \cdot \nabla \psi'(r, t)] + \frac{1}{2m} \left[i \nabla^2 \phi(r, t) \psi'(r, t) + (\nabla \phi(r, t))^2 \psi'(r, t) \right]$$

$$= -\frac{\partial \phi(r, t)}{\partial t} \psi'(r, t) - \frac{\nabla^2}{2m} \psi'(r, t) + V(r) \psi'(r, t) + \frac{i}{m}[\nabla \phi(r, t) \cdot \nabla \psi'(r, t)] + \frac{1}{2m}\left[i \nabla^2 \phi(r, t) \psi'(r, t) + (\nabla \phi(r, t))^2 \psi'(r, t)\right]$$

$$\left(i \frac{\partial}{\partial t} + \frac{\partial \phi(r, t)}{\partial t}\right) \psi'(r, t) = \frac{[-i \nabla - \nabla \phi(r, t)]^2}{2m} \psi'(r, t) + V(r) \psi'(r, t) \tag{3.65}$$

$$\frac{[-i \nabla - \nabla \phi(r, t)]^2}{2m} \psi'(r, t) = \frac{[-i \nabla - \nabla \phi(r, t)] \cdot [-i \nabla - \nabla \phi(r, t)]}{2m} \psi'(r, t) = \frac{[-i \nabla - \nabla \phi(r, t)]}{2m} \cdot [-i \nabla \psi'(r, t) - \psi'(r, t) \nabla \phi(r, t)] =$$

$$-\frac{\nabla^2 \psi'(r, t)}{2m} + \frac{i \nabla}{2m} \cdot [\psi'(r, t) \nabla \phi(r, t)] - \frac{\nabla \phi(r, t)}{2m} \cdot [-i \nabla \psi'(r, t) - \psi'(r, t) \nabla \phi(r, t)] \tag{3.66}$$

$$= -\frac{\nabla^2 \psi'(r, t)}{2m} + \frac{i}{m}[\nabla \psi'(r, t) . \nabla \phi(r, t)] + \frac{i}{2m}\left[\psi'(r, t) \nabla^2 \phi(r, t)\right] + \frac{[\nabla \phi(r, t)]^2}{2m} \psi'(r, t)$$

Q: Can we make the Schrodinger equation invariant under local phase shift?

A: Yes. Using charge particles.

Schrodinger equation for a particle with charge e. Just change the momentum operator $\vec{p}$ into $\vec{p} + e \vec{A} / c$, where $\vec{A}$ is the vector potential, and change $i \partial_t$ into $i \partial_t - e \Phi$, where Φ is the Electric potential. (minimal coupling).

$$(i \hbar \partial_t - e \Phi) \psi(r, t) = \left[\frac{1}{2m}\left(-i \hbar \nabla - \frac{e}{c} \vec{A}\right)^2\right] \psi(r, t) + V(r) \psi(r, t) \tag{3.67}$$

$$\left(i \hbar \partial_t - e \Phi + \hbar \frac{\partial \phi(r, t)}{\partial t}\right) \psi'(r, t) = \frac{1}{2m}\left(-i \hbar \nabla - \frac{e}{c} \vec{A} - \hbar \nabla \phi(r, t)\right)^2 \psi(r, t) + V(r) \psi(r, t) \tag{3.68}$$

If we define:

$$\Phi' = \Phi - \frac{\hbar}{e} \frac{\partial \phi(r, t)}{\partial t} \tag{3.69}$$

and

$$\vec{A}' = \vec{A} + \frac{c \hbar}{e} \nabla \phi(r, t) \tag{3.70}$$

we got:

$$(i \hbar \partial_t - e \Phi') \psi'(r, t) = \frac{1}{2m}\left(-i \hbar \nabla - \frac{e}{c} \vec{A}'\right)^2 \psi(r, t) + V(r) \psi(r, t) \tag{3.71}$$

The SE is invariant under a local phase shift, if we changes the potential and vector potential as defined above. The change of Φ and A is just a gauge transformation. It doesn't change any physics (Same E and B). So we say that the SE for a charged particle is invariant under local phase shift.

Because this local gauge shift is related with a gauge transformation, sometime we just call this local phase shift a a gauge transformation.

In another convention, which is probably more widely used, on absorb the constant c $\hbar$/e into the definition of ϕ, so the SE is invariant under gauge transformation:

$$\Phi \rightarrow \Phi' = \Phi - \frac{\partial \phi(r, t)}{c \partial t} \tag{3.72}$$

$$\vec{A} \rightarrow \vec{A}' = \vec{A} + \nabla \phi(r, t) \tag{3.73}$$

$$\Psi(r, t) \rightarrow \Psi('r, t) = \Psi(r, t) \exp\left(i \frac{e}{c \hbar} \phi\right) \tag{3.74}$$

Bottom line: local phase symmetry is directly related with a gauge field. For the Bloch waves, which as a local phase symmetry in the k-space, we should also think about "gauge fields" but in k-space.

3.6. Berry connection and Berry curvature: "Vector potential" and "magnetic fields" in k-space

3.6.1. Berry connection

Define the Berry connection:

$$\vec{\mathcal{A}}_n = -i \left\langle u_{n,k} \left| \nabla_{\vec{k}} \right| u_{n,k} \right\rangle \tag{3.75}$$

In some literature, there is no minus sign in the definition. In some references, k is the momentum and in some others k is the wave vector (they are different by a factor of $\hbar$). Here, I choose k to be the wave vector.

This is a gauge like field in the k-space, pretty similar to the vector potential $\vec{A}$.

Under a local phase shift:

$$\left| u_{n,k} \right\rangle \rightarrow e^{i\,\phi_n(k)} \left| u_{n,k} \right\rangle \tag{3.76}$$

$$\vec{\mathcal{A}}_n \rightarrow \vec{\mathcal{A}}_n{}' = -i \left\langle u_{n,k} \left| e^{-i\,\phi_n(k)} \nabla_{\vec{k}} e^{i\,\phi_n(k)} \right| u_{n,k} \right\rangle = -i \left\langle u_{n,k} \left| \nabla_{\vec{k}} \right| u_{n,k} \right\rangle + \nabla_{\vec{k}} \phi_n(k) \left\langle u_{n,k} \mid u_{n,k} \right\rangle = \vec{\mathcal{A}}_n + \nabla_{\vec{k}} \phi_n(k) \tag{3.77}$$

$\vec{\mathcal{A}}_n$ changes like *a* gauge field (the vector potential). But $\vec{\mathcal{A}}_n$ lives in the $k-$ space and all derivatives are derivatives of k, instead of r.

3.6.2. Berry curvature

We know that the vector potential is not a physical observable, and its value depends on the gauge choice. The quantity with physical meaning is the curl of it, which is the magnetic field B. Here, it is the same story.

$$\mathcal{F}_n = \nabla_k \times \vec{\mathcal{A}}_n = -i \,\epsilon_{ij}\, \partial_{k_i} \left\langle u_{n,k} \left| \partial_{k_j} \right| u_{n,k} \right\rangle = -i \,\epsilon_{ij} \left\langle \partial_{k_i} u_{n,k} \left| \partial_{k_j} u_{n,k} \right\rangle \right. \tag{3.78}$$

3.6.3. The position operator in a lattice
We know that without lattice, $p = -i\,\hbar\,\nabla_r$ and $r = i\,\hbar\,\nabla_p$. What will happen if we have a lattice?

In the continuum we know that $p = -i\,\hbar\,\nabla_r$ and $r = i\,\hbar\,\nabla_p$. To see this, we simply use the plan wave expansions and write any wave-functions in a superposition of plan waves.

$$\Psi\!\left(\vec{k}\right) = \int d\,\vec{r}\, \Psi\!\left(\vec{r}\right) e^{-i\,\vec{k}\cdot\vec{r}} \tag{3.79}$$

$$\Psi\!\left(\vec{r}\right) = \int d\,\vec{k}\, \Psi\!\left(\vec{k}\right) e^{i\,\vec{k}\cdot\vec{r}} \tag{3.80}$$

To prove $r = i\,\nabla_p$, the most straight-forward way is to use

$$\Psi\!\left(\vec{k}\right) = \int d\,\vec{r}\, \Psi\!\left(\vec{r}\right) e^{-i\,\vec{k}\cdot\vec{r}} \tag{3.81}$$

$$\vec{r}\,\Psi\!\left(\vec{k}\right) = \int d\,\vec{r}\,\vec{r}\, \Psi\!\left(\vec{r}\right) e^{-i\,\vec{k}\cdot\vec{r}} = \int d\,\vec{r}\, \Psi\!\left(\vec{r}\right)\!\left[i\,\vec{\partial}_k\, e^{-i\,\vec{k}\cdot\vec{r}}\right] = i\,\vec{\partial}_k \int d\,\vec{r}\, \Psi\!\left(\vec{r}\right) e^{-i\,\vec{k}\cdot\vec{r}} = i\,\vec{\partial}_k\, \Psi\!\left(\vec{k}\right) \tag{3.82}$$

Here, we will show a different (less straightforward approach), which can be generalized to Bloch waves easily using the inverse transformation:

$$\Psi\!\left(\vec{r}\right) = \int d\,\vec{k}\, \Psi\!\left(\vec{k}\right) e^{i\,\vec{k}\cdot\vec{r}} \tag{3.83}$$

Here we ask the question, what operator $\vec{X}$ can we use to make the two sides of the equation equal. Because we have an position operator on the l.h.s., this means that $\vec{X}$ is the position operator.

$$\vec{r}\,\Psi\!\left(\vec{r}\right) = \int d\,\vec{k}\,\vec{X}\, \Psi\!\left(\vec{k}\right) e^{i\,\vec{k}\cdot\vec{r}} \tag{3.84}$$

$$\vec{r}\,\Psi\left(\vec{r}\right) = \int d\vec{k}\,\vec{r}\,\Psi\left(\vec{k}\right)e^{i\vec{k}\cdot\vec{r}} = \int d\vec{k}\,\Psi\left(\vec{k}\right)\left[-i\,\partial_k\,e^{i\vec{k}\cdot\vec{r}}\right] = \int d\vec{k}\left[i\,\partial_k\,\Psi\left(\vec{k}\right)\right]e^{i\vec{k}\cdot\vec{r}} \tag{3.85}$$

So we have $\vec{X} = i\,\nabla_k$

For Bloch waves, we can do the same thing. In general, any wavefunction can be written as a superposition of the Bloch waves:

$$\Psi\left(\vec{r}\right) = \sum_n \int d\vec{k}\,\Psi_n\left(\vec{k}\right)\psi_{n,k}\left(\vec{r}\right) = \sum_n \int d\vec{k}\,\Psi_n\left(\vec{k}\right)u_{n,k}\left(\vec{r}\right)e^{i\vec{k}\cdot\vec{r}} \tag{3.86}$$

This is similar to a Fourier transform, but we use Bloch waves as basis instead of the planes. Here the inverse transformation is not so straightforward to write down.

If I want to know the position of this wavefunction:$\Psi(r)$, we simply use $\vec{r}\,\Psi\left(\vec{r}\right)$, and now we ask what operator $\vec{X}$ can we use to satisfies:

$$\vec{r}\,\Psi\left(\vec{r}\right) = \sum_n \int d\vec{k}\,\vec{X}\,\Psi_n\left(\vec{k}\right)\psi_{n,k}\left(\vec{r}\right) \tag{3.87}$$

This $\vec{X}$ operator is the position operator in k-space

$$\vec{r}\,\Psi\left(\vec{r}\right) = \vec{r}\sum_n \int d\vec{k}\,\Psi_n\left(\vec{k}\right)\psi_{n,k}\left(\vec{r}\right) = \sum_n \int d\vec{k}\,\Psi_n\left(\vec{k}\right)u_{n,k}\left(\vec{r}\right)\vec{r}\,e^{i\vec{k}\cdot\vec{r}} = \sum_n \int d\vec{k}\,\Psi_n\left(\vec{k}\right)u_{n,k}\left(\vec{r}\right)\left(-i\,\partial_k\,e^{i\vec{k}\cdot\vec{r}}\right) =$$
$$\sum_n \int d\vec{k}\left[i\,\partial_k\,\Psi_n\left(\vec{k}\right)u_{n,k}\left(\vec{r}\right)\right]e^{i\vec{k}\cdot\vec{r}} = \sum_n \int d\vec{k}\left[i\,\partial_k\,\Psi_n\left(\vec{k}\right)u_{n,k}\left(\vec{r}\right) + \Psi_n\left(\vec{k}\right)i\,\partial_k\,u_{n,k}\left(\vec{r}\right)\right]e^{i\vec{k}\cdot\vec{r}} \tag{3.88}$$

$$\vec{r}\,\Psi\left(\vec{r}\right) = \sum_n \int d\vec{k}\left[i\,\partial_k\,\Psi_n\left(\vec{k}\right)\psi_{n,k}\left(\vec{r}\right) + \Psi_n\left(\vec{k}\right)\int dr'\,\delta(r-r')\,i\,\partial_k\,u_{n\,k}\left(\vec{r'}\right)e^{i\vec{k}\cdot\vec{r}}\right]$$
$$= \sum_n \int d\vec{k}\left[i\,\partial_k\,\Psi_n\left(\vec{k}\right)\psi_{n,k}\left(\vec{r}\right) + \Psi_n\left(\vec{k}\right)\int dr'\sum_m u^*_{m,k}(r')\,u_{m,k}(r)\,i\,\partial_k\,u_{n\,k}\left(\vec{r'}\right)e^{i\vec{k}\cdot\vec{r}}\right]$$
$$= \sum_n \int d\vec{k}\left\{i\,\partial_k\,\Psi_n\left(\vec{k}\right)\psi_{n,k}\left(\vec{r}\right) + \Psi_n\left(\vec{k}\right)\sum_m u_{m,k}(r)\left[\int dr'\,u^*_{m,k}(r')\,i\,\partial_k\,u_{n\,k}\left(\vec{r'}\right)\right]e^{i\vec{k}\cdot\vec{r}}\right\} \tag{3.89}$$
$$= \sum_n \int d\vec{k}\left\{i\,\partial_k\,\Psi_n\left(\vec{k}\right)\psi_{n,k}\left(\vec{r}\right) + \Psi_n\left(\vec{k}\right)\sum_m u_{m,k}(r)[-\mathcal{A}_{mn}(k)]\,e^{i\vec{k}\cdot\vec{r}}\right\}$$
$$= \sum_n \left[\int d\vec{k}\,i\,\partial_k\,\Psi_n\left(\vec{k}\right)\psi_{n,k}\left(\vec{r}\right) - \sum_m \vec{\mathcal{A}}_{m,n}\,\Psi_n\left(\vec{k}\right)\psi_{m,k}\left(\vec{r}\right)\right]$$

If one only focus on a single band, say band n, we have

$$\vec{r}\,\Psi\left(\vec{r}\right) = \int d\vec{k}\left[i\,\partial_k\,\Psi_n\left(\vec{k}\right)\psi_{n,k}\left(\vec{r}\right) - \int d\vec{r}\,\vec{\mathcal{A}}_{n,n}\,\Psi_n\left(\vec{r}\right)\psi_{n,k}\left(\vec{r}\right)\right] \tag{3.90}$$

So:

$$\vec{r} = i\,\vec{\partial}_k - \vec{\mathcal{A}}_n \tag{3.91}$$

Just like what we are very familiar with for charged particles.

$$\vec{p} = -i\,\nabla_x - \frac{e}{c}\,\vec{A} \tag{3.92}$$

More generic formula:

$$\vec{r} = i\,\vec{\partial}_k\,\delta_{n,m} - \vec{\mathcal{A}}_{m.n} \tag{3.93}$$

But we can usually ignore the terms $m \neq n$ and they don't contribute to the Hall effect.

It is also worthwhile to notice that $\vec{r} = i\,\nabla_k - \vec{\mathcal{A}}_n$ is invariant under local phase change in the k-space $|u_{n,k}\rangle \to e^{i\,\phi_n(k)}\,|u_{n,k}\rangle$.

For a generic wavefunction, one can write as superpositions of Bloch waves

$$\Psi\left(\vec{r}\right) = \sum_n \int d\vec{k}\,\Psi_n\left(\vec{k}\right)\psi_{n,k}\left(\vec{r}\right) \tag{3.94}$$

The coefficent $\Psi_n\left(\vec{k}\right)$ is our wave-function. If we act $\left(i\,\nabla_k - \vec{\mathcal{A}}_n\right)$ on to $\Psi_n\left(\vec{k}\right)$, and do the gauge transformation, one gets

$$\left(i \, \nabla_k - \vec{\mathcal{A}}_n \right) \Psi_n\left(\vec{k} \right) \to \left(i \, \nabla_k - \vec{\mathcal{A}}_n - \nabla_{\vec{k}} \, \phi_n(k) \right) e^{-i \, \phi_n(k)} \, \Psi_n\left(\vec{k} \right) \tag{3.95}$$

Here we use the fact that under the gauge transformation, $\psi_{n,\vec{k}}\left(\vec{r} \right) \to e^{-i\,\phi_n(k)} \Psi_n\left(\vec{k} \right)$. To keep $\Psi\left(\vec{r} \right)$ invariant, $\Psi_n\left(\vec{k} \right) \to e^{-i\,\phi_n(k)} \Psi_n\left(\vec{k} \right)$

$$\left(i \, \nabla_k - \vec{\mathcal{A}}_n \right) \psi_{n,k}\left(\vec{r} \right) \to \left(i \, \nabla_k - \vec{\mathcal{A}}_n - \nabla_{\vec{k}} \, \phi_n(k) \right) e^{-i \, \phi_n(k)} \, \Psi_n\left(\vec{k} \right) =$$
$$e^{i\,\phi_n(k)} \left(i \, \nabla_k - \vec{\mathcal{A}}_n - \nabla_{\vec{k}} \, \phi_n(k) \right) \Psi_n\left(\vec{k} \right) + \nabla_k \, \phi_n(k) \, e^{i\,\phi_n(k)} \, \Psi_n\left(\vec{k} \right) = e^{i\,\phi_n(k)} \left(i \, \nabla_k - \vec{\mathcal{A}}_n \right) \Psi_n\left(\vec{k} \right) \tag{3.96}$$

$$\Psi^*_n\left(\vec{k} \right) \left(i \, \nabla_k - \vec{\mathcal{A}}_n \right) \Psi_n\left(\vec{k} \right) \to e^{-i\,\phi_n(k)} \, \Psi^*_n\left(\vec{k} \right) e^{i\,\phi_n(k)} \left(i \, \nabla_k - \vec{\mathcal{A}}_n \right) \Psi_n\left(\vec{k} \right) = \Psi^*_n\left(\vec{k} \right) \left(i \, \nabla_k - \vec{\mathcal{A}}_n \right) \Psi_n\left(\vec{k} \right) \tag{3.97}$$

On the other hand, neither $i \, \nabla_k$ nor $\vec{\mathcal{A}}_n$ is invariant under this local phase change. So their value depends on the choice of gauge (choice of phases). This means that they are NOT physical observables and their value has no physics meaning. Only the sum of they two is meaningful, which is invariant under this "gauge transformation".

3.6.4. Equations of motion and anomalous velocity

We know that under electric and magnetic fields (E and B), the equations of motion for a particle with charge e is

$$\frac{d\vec{p}}{dt} = \vec{F} = e\,\vec{E} + \frac{e}{c}\,\vec{v} \times \vec{B} = -e\,\vec{\nabla}_r \phi(r) + \frac{e}{c}\,\frac{d\vec{r}}{dt} \times \left[\vec{\nabla}_r \times \vec{A}(r) \right] \tag{3.98}$$

where ϕ is the electric potential and A is the vector potential

In quantum mechanics, p and r are conjugate variables. We can choose to write the Hamiltonian as a function of r and ∂_r, or p and ∂_p. These two descriptions are equivalent. Since we can freely choose to use r or p, a natural question to ask is:

Q: What is the equations of motion for $\vec{r}$? Is that EOM similar to the one shown above? Are there something like to ϕ and $\vec{A}$ in the E.O.M. for $\vec{r}$

A: The EOM of $\vec{r}$ looks almost identical to $\vec{p}$. We just need to switch r and p, and use the dispersion relation $\epsilon\left(\vec{p} \right)$ to replace $\phi\left(\vec{r} \right)$ and the Berry connection $\vec{\mathcal{A}}(p)$ to replace the vector potential $\vec{A}\left(\vec{r} \right)$

$$\frac{d\vec{r}}{dt} = \nabla_p \epsilon(p) + \hbar\,\frac{d\vec{p}}{dt} \times \left[\vec{\nabla}_p \times \vec{\mathcal{A}}(p) \right] \tag{3.99}$$

For the special case of a free fermion with no E and B fields: $\epsilon = p^2 / 2\,m$ and $\vec{\mathcal{A}}(p)=0$, this equation turns into

$$\frac{d\vec{r}}{dt} = \frac{\nabla_p \, p^2}{2\,m} = \frac{2\,\vec{p}}{2\,m} = \vec{v} \tag{3.100}$$

In many cases, $\nabla_p \epsilon(p)$ is the definition of the "velocity" for a quantum particle. However, if $\vec{\mathcal{A}}(p)$ is nonzero. This velocity is not just $\nabla_p \epsilon(p)$. There is an extra term $\frac{d\vec{p}}{dt} \times \left[\vec{\nabla}_p \times \vec{\mathcal{A}}(p) \right]$, which is known as the "anomalous velocity".

Let's first go back to the EOM of $\vec{p}$. And demonstrate how this equation can be obtain in quantum mechanics.

the Hamiltonian is in general a function of r and p. In the real space, it is a function of $\vec{r}$ and $\vec{\partial}_r$.

$$H = H\left(\vec{r}, \, \vec{p} \right) = H\left(\vec{r}, \, -i\,\hbar\,\vec{\partial}_r \right) \tag{3.101}$$

For charged particle, the momentum $\vec{p}$ changes into $\vec{p} = -i\,\hbar\,\vec{\partial}_r - \frac{e}{c}\,\vec{A}(r)$. So the Hamiltonian is a function of $\vec{r}$ and

$$H = H\left(\vec{r}, \, -i\,\hbar\,\vec{\partial}_r - \frac{e}{c}\,\vec{A}(r) \right) \tag{3.102}$$

In many cases, we can separate the Hamiltonian into a kinetic energy part and a potential energy part:

$$H = K\left(\vec{p}\right) + e\,\phi\left(\vec{r}\right) = K\left[-i\,\hbar\,\vec{\partial}_r - \frac{e}{c}\,\vec{A}(r)\right] + e\,\phi\left(\vec{r}\right) \tag{3.103}$$

Here K is the kinetic energy, which only depends on the momentum p. For charged particle, $\vec{p} = -i\,\hbar\,\vec{\partial}_r - \frac{e}{c}\,\vec{A}(r)$. The second term is the electric

energy (potential energy), which only depends on $\vec{r}$.

Now, momentum $\vec{p} = -i\,\hbar\,\vec{\partial}_r - \frac{e}{c}\,\vec{A}(r)$, so

$$\frac{d\,p_i}{d\,t} = \frac{i}{\hbar}[H,\,p_i] = \frac{i}{\hbar}\left[H,\,-i\,\hbar\,\partial_i - \frac{e}{c}\,A_i(r)\right] = [H,\,\partial_i] - \frac{i}{\hbar}\frac{e}{c}[H,\,A_i(r)] =$$

$$-\partial_{r_i} H - \frac{e}{c}\frac{d}{d\,t}A_i\left(\vec{r}\right) = -\partial_i\,K\left(\vec{p}\right) - e\,\partial_i\,\phi\left(\vec{r}\right) - \frac{e}{c}\frac{\partial A_i}{\partial r_j}\frac{d\,r_j}{d\,t} = -\frac{\partial K}{\partial p_j}\frac{\partial p_j}{\partial r_i} - e\,\partial_i\,\phi\left(\vec{r}\right) - \frac{e}{c}\frac{\partial A_i}{\partial r_j}\frac{d\,r_j}{d\,t} =$$

$$\frac{e}{c}\frac{\partial K}{\partial p_j}\frac{\partial A_j}{\partial r_i} - e\,\partial_i\,\phi\left(\vec{r}\right) - \frac{e}{c}\frac{\partial A_i}{\partial r_j}\frac{d\,r_j}{d\,t} = \frac{e}{c}\frac{\partial H}{\partial p_j}\frac{\partial A_j}{\partial r_i} - e\,\partial_i\,\phi\left(\vec{r}\right) - \frac{e}{c}\frac{\partial A_i}{\partial r_j}\frac{d\,r_j}{d\,t} = \tag{3.104}$$

$$\frac{e}{c}\frac{i}{\hbar}[H,\,r_j]\frac{\partial A_j}{\partial r_i} - e\,\partial_i\,\phi\left(\vec{r}\right) - \frac{e}{c}\frac{\partial A_i}{\partial r_j}\frac{d\,r_j}{d\,t} = \frac{e}{c}\frac{d\,r_j}{d\,t}\frac{\partial A_j}{\partial r_i} - e\,\partial_i\,\phi\left(\vec{r}\right) - \frac{e}{c}\frac{\partial A_i}{\partial r_j}\frac{d\,r_j}{d\,t} = -e\,\partial_i\,\phi\left(\vec{r}\right)\frac{d\,r_j}{d\,t} +$$

$$\frac{e}{c}\frac{d\,r_j}{d\,t}\left(\frac{\partial A_j}{\partial r_i} - \frac{\partial A_i}{\partial r_j}\right) = -e\,\partial_i\,\phi\left(\vec{r}\right)\frac{d\,r_j}{d\,t} + \frac{e}{c}\,v_j\left(\frac{\partial A_j}{\partial r_i} - \frac{\partial A_i}{\partial r_j}\right) = -e\,\partial_i\,\phi\left(\vec{r}\right) + \frac{e}{c}\left\{\frac{d\,\vec{r}}{d\,t}\times\left[\vec{\nabla}_r \times \vec{A}(r)\right]\right\}_i$$

Now, let's go to the k-space

In a band theory, the Bloch waves are eigen-states of the Hamiltonian with dispersion relation $\epsilon_n(k)$. For simplicity, here we focus on one of the energy bands, so the Hamiltonian is

$$H = \epsilon_n\left(\vec{p}\right) \tag{3.105}$$

For charged particles, $\vec{p} \to \vec{p} - \frac{e}{c}\,\vec{A}\left(\vec{r}\right)$

$$H = \epsilon_n\left(\vec{p} - \frac{e}{c}\,\vec{A}\left(\vec{r}\right)\right) \tag{3.106}$$

So H is in general a function of $\vec{p}$ and $\vec{r}$.

We know that for Bloch waves: $\vec{r} = i\,\hbar\,\nabla_p - \vec{\mathcal{A}}_n$. So

$$H = \mathcal{H}\left(\vec{p},\,\vec{r}\right) = \mathcal{H}\left(\vec{p},\,i\,\hbar\,\nabla_p - \vec{\mathcal{A}}_n\right) \tag{3.107}$$

This is almost the same as what we have above. So, if we follow the same steps, we find exactly the same equations of motion for $\vec{r}$.

$$\frac{d\,r_i}{d\,t} = \frac{i}{\hbar}[H,\,r_i] = \frac{i}{\hbar}\left[H,\,i\,\hbar\,\partial_{\mathrm{pi}} - \mathcal{A}_i(p)\right] =$$

$$-[H,\,\partial_{\mathrm{pi}}] - \frac{i}{\hbar}[H,\,\mathcal{A}_i(p)] = \frac{d}{d\,p_i}\mathcal{H}\left(\vec{p},\,\vec{r}\right) - \frac{i}{\hbar}\left[\mathcal{H}\left(\vec{p},\,\vec{r}\right),\,\mathcal{A}_i\left(\vec{p}\right)\right] = \partial_{p_i}\mathcal{H}\left(\vec{p},\,\vec{r}\right) + \frac{\partial\mathcal{H}\left(\vec{p},\,\vec{r}\right)}{\partial r_j}\frac{\partial r_j}{\partial p_i} - \frac{d\,\mathcal{A}_i\left(\vec{p}\right)}{d\,t} = \tag{3.108}$$

$$\partial_{p_i}\mathcal{H}\left(\vec{p},\,\vec{r}\right) - \frac{\partial\mathcal{H}\left(\vec{p},\,\vec{r}\right)}{\partial r_j}\frac{\partial\mathcal{A}_j}{\partial p_i} - \frac{\partial\mathcal{A}_i\left(\vec{p}\right)}{\partial p_j}\frac{d\,p_j}{d\,t} = \partial_{p_i}\mathcal{H}\left(\vec{p},\,\vec{r}\right) + \frac{d\,p_j}{d\,t}\frac{\partial\mathcal{A}_j}{\partial p_i} - \frac{\partial\mathcal{A}_i}{\partial p_j}\frac{d\,p_j}{d\,t}$$

Here we used that

$$\frac{\partial\mathcal{H}\left(\vec{p},\,\vec{r}\right)}{\partial r_j} = -\frac{i}{\hbar}\left[\mathcal{H}\left(\vec{p},\,\vec{r}\right),\,p_j\right] = -\frac{d\,p_j}{d\,t} \tag{3.109}$$

So

$$\frac{d\,r_i}{d\,t} = \partial_{p_i}\,\mathcal{H}\left(\vec{k},\,\vec{r}\right) + \frac{d\,p_j}{d\,t}\left(\frac{\partial \mathcal{A}_j}{\partial\,p_i} - \frac{\partial \mathcal{A}_i}{\partial\,p_j}\right) \tag{3.110}$$

In the weak field limit,

$$H = \epsilon_n\!\left(\vec{p} - \frac{e}{c}\,\vec{A}\!\left(\vec{r}\right)\right) \approx \epsilon_n(p) \tag{3.111}$$

So

$$\frac{d\,r_i}{d\,t} = \partial_{p_i}\,\mathcal{H}\left(\vec{p},\,\vec{r}\right) + \frac{d\,p_j}{d\,t}\left(\frac{\partial \mathcal{A}_j}{\partial\,p_i} - \frac{\partial \mathcal{A}_i}{\partial\,p_j}\right) \approx \partial_{p_i}\,\epsilon_n + \frac{d\,p_j}{d\,t}\left(\frac{\partial \mathcal{A}_j}{\partial\,p_i} - \frac{\partial \mathcal{A}_i}{\partial\,p_j}\right) \tag{3.112}$$

$$\frac{d\,\vec{r}}{d\,t} = \nabla_p\,\epsilon(p) + \frac{d\,\vec{p}}{d\,t} \times \left[\vec{\nabla}_p \times \vec{\mathcal{A}}(p)\right] \tag{3.113}$$

3.6.5. Berry curvature and the Hall effect

Reference: Haldane, Berry Curvature on the Fermi Surface: Anomalous Hall Effect as a Topological Fermi-Liquid Property, Phys. Rev. Lett. 93, 206602 (2004).

In the presence of E and B fields, the Newton's second law reads

$$\frac{d\,\vec{p}}{d\,t} = \vec{F} = e\,\vec{E} + e\,\vec{v}\times\vec{B} \tag{3.114}$$

$$\frac{d\,\vec{r}}{d\,t} = \nabla_p\,\epsilon(p) + \frac{d\,\vec{p}}{d\,t} \times \left[\vec{\nabla}_p \times \vec{\mathcal{A}}(p)\right] \tag{3.115}$$

If all the electrons have the same velocity, we have current:

$$j = e\,n\,v = \frac{e\,N\,v}{A} \tag{3.116}$$

If the velocities are different,

$$j = \frac{e}{A}\sum_m v_m = \frac{e}{A}\sum_m \frac{d\,\vec{r_m}}{d\,t} = e\sum_{n\ \text{fully occupied bands}} \frac{A}{h^2}\int_{\text{BZ}} d\,\vec{p}\left\{\nabla_p\,\epsilon_n(p) + \frac{d\,\vec{p}}{d\,t}\times\left[\vec{\nabla}_p \times \vec{\mathcal{A}}_n(p)\right]\right\} +$$

$$e\sum_{n\ \text{partially filled bands}} \frac{A}{\hbar^2}\int_{\text{occupied part}} d\,\vec{p}\left\{\nabla_p\,\epsilon_n(p) + \frac{d\,\vec{p}}{d\,t}\times\left[\vec{\nabla}_p \times \vec{\mathcal{A}}_n(p)\right]\right\} \tag{3.117}$$

For an insulator, we only have fully filled bands so we can drop the last term.

$$\frac{d\,\vec{r}}{d\,t} = \nabla_p\,\epsilon_n(p) + \frac{d\,\vec{p}}{d\,t} \times \left[\vec{\nabla}_p \times \vec{\mathcal{A}}_n(p)\right] = \nabla_p\,\epsilon_n(p) + \left(e\,\vec{E} + e\,\vec{v}\times\vec{B}\right)\times\left[\vec{\nabla}_p \times \vec{\mathcal{A}}_n(p)\right] \tag{3.118}$$

Here, one term is important:

$$e\,\vec{E}\times\left[\vec{\nabla}_p \times \vec{\mathcal{A}}_n(p)\right] \tag{3.119}$$

If we changes all the p back to k, we get one extra $\hbar$

$$e\,\vec{E}\times\left[\vec{\nabla}_p \times \vec{\mathcal{A}}_n(p)\right] = \frac{1}{\hbar}\,e\,\vec{E}\times\left[\vec{\nabla}_k \times \vec{\mathcal{A}}_n(k)\right] \tag{3.120}$$

In addition, $\int d\text{p}$ turns into $\hbar^2 \int d\,k$

This one gives Hall effect: velocity of the electron is perpendicular to the direction of $\vec{E}$. So if we only care about Hall conductivity, the Hall current is

$$j_H = \frac{e^2}{2\pi h} \sum_{n \text{ fully occupied bands}} \int_{\text{BZ}} d\vec{k}\, \vec{E} \times \left[\vec{\nabla}_k \times \vec{\mathcal{A}}_n(k) \right] = \vec{E} \times \left[\frac{e^2}{h}\frac{1}{2\pi} \sum_{n \text{ fully occupied bands}} \int_{\text{BZ}} d\vec{k}\, \vec{E} \times \vec{\mathcal{F}}_n(k) \right] \tag{3.121}$$

$$\sigma_{\text{xy}} = \frac{e^2}{h} \sum_{n \text{ fully occupied bands}} \left[\frac{1}{2\pi} \int_{\text{BZ}} d\vec{k}\, \vec{\mathcal{F}}_n(k) \right] \tag{3.122}$$

In the next section, we will show that the integral here is always an integer, and it is a topological index.

From now on, I will use the theorist's unit: $k_B = e = c = h = \hbar = 2\pi = 1$ so

$$\sigma_{\text{xy}} = \sum_n \left[-\frac{i}{2\pi} \int_{\text{BZ}} d\vec{k}\, \epsilon_{\text{ij}} \left\langle \partial_{\text{ki}} u_{n,k} \mid \partial_{\text{kj}} u_{n,k} \right\rangle \right] \tag{3.123}$$

is just an integer.

3.7. Dirac Quantization, Gauss– Bonnet theorem and the TKNN (Thouless —Kohmoto —Nightingale —den Nijs) Invariant

From the mathematical point of the view, the following three objects are the same thing (fiber bundles): the magnetic field B, the Berry curvature $\mathcal{F}$, and the Gaussian curvature of K (geometry). All of them are described by the same mathematical structure: fiber bundles.

In the next a few sections, we will investigate the integral of B, $\mathcal{F}$ and K on a closed 2D manifold. (close: no boundary). And shows that they are all quantized due to topological reasons, which is known as topological quantization.

$$\oiint_M B \cdot dS = \oiint_M B_n \, dS = \frac{c\,\hbar}{2\,q_e}\, n \qquad \text{quantized}: \ n \text{ is an integer,}$$

known as the magnetic charge,   which measures the number of magnetic monopole inside M $\tag{3.124}$

$$\oiint_M K \, dS = 2\pi\,\chi_M \qquad\qquad \text{quantized}: \ \chi_M \text{ is an even integer,}$$

known as the Euler characteristic,   which measures the topological nature of the manifold M $\tag{3.125}$

$$\oiint_{\text{BZ}} \mathcal{F} \, d\vec{k} = 2\pi\,C \qquad\qquad \text{quantized}: \ C \text{ is an integer,}$$

known as the TKNN invariant or the Chern number,   which measures the quantized Hall conductivity for *a* topological insulator $\tag{3.126}$

3.8. Magnetic monopole and Dirac quantization condition

Reference: M Nakahara, Geometry, topology and physics, IOP

For electric charge, the Gauss's law tell us that

$$q_e = \oiint_M E \cdot dS \tag{3.127}$$

In Maxwell's equations, this is:

$$\nabla \cdot E = \rho \tag{3.128}$$

For magnetic fields, we can do the same thing:

$$q_m = \oiint_M B \cdot dS \tag{3.129}$$

Without magnetic monopoles, $q_m = 0$, because $\nabla \cdot B = 0$

Assume that there is a magnetic monopole with charge q_m, what will happen?

The *B* fields is:

$$j_H = \frac{e^2}{2\pi h} \sum_{n \text{ fully occupied bands}} \int_{BZ} d\vec{k}\, \vec{E} \times \left[\vec{\nabla}_k \times \vec{\mathcal{A}}_n(k) \right] = \vec{E} \times \left[\frac{e^2}{h}\, \frac{1}{2\pi} \sum_{n \text{ fully occupied bands}} \int_{BZ} d\vec{k}\, \vec{E} \times \vec{\mathcal{F}}_n(k) \right] \tag{3.121}$$

$$\sigma_{xy} = \frac{e^2}{h} \sum_{n \text{ fully occupied bands}} \left[\frac{1}{2\pi} \int_{BZ} d\vec{k}\, \vec{\mathcal{F}}_n(k) \right] \tag{3.122}$$

In the next section, we will show that the integral here is always an integer, and it is a topological index.

From now on, I will use the theorist's unit: $k_B = e = c = h = \hbar = 2\pi = 1$ so

$$\sigma_{xy} = \sum_n \left[-\frac{i}{2\pi} \int_{BZ} d\vec{k}\, \epsilon_{ij} \left\langle \partial_{ki} u_{n,k} \mid \partial_{kj} u_{n,k} \right\rangle \right] \tag{3.123}$$

is just an integer.

3.7. Dirac Quantization, Gauss– Bonnet theorem and the TKNN (Thouless —Kohmoto —Nightingale —den Nijs) Invariant

From the mathematical point of the view, the following three objects are the same thing (fiber bundles): the magnetic field B, the Berry curvature $\mathcal{F}$, and the Gaussian curvature of K (geometry). All of them are described by the same mathematical structure: fiber bundles.

In the next a few sections, we will investigate the integral of B, $\mathcal{F}$ and K on a closed 2D manifold. (close: no boundary). And shows that they are all quantized due to topological reasons, which is known as topological quantization.

$$\oiint_M B \cdot dS = \oiint_M B_n\, dS = \frac{c\hbar}{2 q_e} n \qquad \text{quantized}:\ n \text{ is an integer}, \tag{3.124}$$

known as the magnetic charge, which measures the number of magnetic monopole inside M

$$\oiint_M K\, dS = 2\pi \chi_M \qquad \text{quantized}:\ \chi_M \text{ is an even integer}, \tag{3.125}$$

known as the Euler characteristic, which measures the topological nature of the manifold M

$$\oiint_{BZ} \mathcal{F}\, d\vec{k} = 2\pi C \qquad \text{quantized}:\ C \text{ is an integer}, \tag{3.126}$$

known as the TKNN invariant or the Chern number, which measures the quantized Hall conductivity for *a* topological insulator

3.8. Magnetic monopole and Dirac quantization condition

Reference: M Nakahara, Geometry, topology and physics, IOP

For electric charge, the Gauss's law tell us that

$$q_e = \oiint_M E \cdot dS \tag{3.127}$$

In Maxwell's equations, this is:

$$\nabla \cdot E = \rho \tag{3.128}$$

For magnetic fields, we can do the same thing:

$$q_m = \oiint_M B \cdot dS \tag{3.129}$$

Without magnetic monopoles, $q_m = 0$, because $\nabla \cdot B = 0$

Assume that there is a magnetic monopole with charge q_m, what will happen?

The *B* fields is:

$$\vec{B} = q_m \frac{\vec{e_r}}{r^2} = q_m \frac{\vec{r}}{r^3} = q_m \frac{(x,\ y,\ z)}{(x^2 + y^2 + z^2)^{3/2}} \tag{3.130}$$

What about the vector potential? $\vec{A} = ?$

$$\nabla \times \vec{A} = \vec{B} \tag{3.131}$$

The value of $\vec{A}$ is not unique, but they are all connected by a gauge transformation.

$$\vec{A} = q_m \frac{(y,\ -x,\ 0)}{r(r-z)} \tag{3.132}$$

$$\nabla \times \vec{A} = q_m \nabla \times \frac{(y,\ -x,\ 0)}{r(r-z)} =$$

$$q_m (\partial_x,\ \partial_y,\ \partial_z) \times \frac{(y,\ -x,\ 0)}{r(r-z)} = q_m \left(\partial_y \frac{0}{r(r-z)} - \partial_z \frac{-x}{r(r-z)},\ \partial_z \frac{y}{r(r-z)} - \partial_x \frac{0}{r(r-z)},\ \partial_x \frac{-x}{r(r-z)} - \partial_y \frac{y}{r(r-z)} \right) = \tag{3.133}$$

$$q_m \left(-\partial_z \frac{-x}{r(r-z)},\ \partial_z \frac{y}{r(r-z)},\ \partial_x \frac{-x}{r(r-z)} - \partial_y \frac{y}{r(r-z)} \right) = q_m \frac{(x,\ y,\ z)}{(x^2 + y^2 + z^2)^{3/2}}$$

This vector potential has one problem. It is singular at the north pole ($z = r$). In fact, one can prove that no matter which gauge one uses, there will always be a singularity point.

Here, I must emphasize that this singularity is not a physical singularity. All physical observables are smooth and non-singular functions at this point. Only A (which is not a measurable quantity) shows singular behavior. In addition, the location of this singularity point is gauge dependent.

For example, we can use another gauge $\vec{A}$, which gives exactly the same B field.

$$\vec{A} = q_m \frac{(-y,\ x,\ 0)}{r(r+z)} \tag{3.134}$$

This A is singular at the south pole $z = -r$.

So, we can use the first one to describe the south hemisphere and the second one to describe the north hemisphere:

$$\vec{A}_N = q_m \frac{(-y,\ x,\ 0)}{r(r+z)} \tag{3.135}$$

$$\vec{A}_S = q_m \frac{(y,\ -x,\ 0)}{r(r-z)} \tag{3.136}$$

At the equator, the vector potential is multivalued (depending on whether we use A_N or A_S), but this is not a problem. Because we know that A is not a physical observable and it IS multivalued. As long as they differ by a gauge transformation, they describe the same physics (same B field). This is indeed the case here: the gauge transformation between A_N and A_S is:

$$\vec{A}_N = \vec{A}_S + 2\,q_m \frac{(-y,\ x,\ 0)}{(r-z)(r+z)} \tag{3.137}$$

At equator: z=0

$$\vec{A}_N = \vec{A}_S + 2\,q_m \frac{(-y,\ x,\ 0)}{r^2} = \vec{A}_N + 2\,q_m \nabla \varphi \tag{3.138}$$

$$\Phi \rightarrow \Phi' = \Phi - \frac{\partial \Lambda(r,\ t)}{c\,\partial t} \tag{3.139}$$

$$\vec{A} \rightarrow \vec{A}' = \vec{A} + \nabla \Lambda(r,\ t) \tag{3.140}$$

$$\Psi(r,\ t) \rightarrow \Psi'(r,\ t) = \Psi(r,\ t) \exp\left(i \frac{q_e}{c\,\hbar} \Lambda \right) \tag{3.141}$$

Here

$$\Lambda(r, t) = 2 \, q_m \, \varphi \tag{3.142}$$

$$\Psi_N(r, t) = \Psi_S(r, t) \exp\left(i \, \frac{q_e}{c \, \hbar} \, \Lambda\right) = \Psi_S(r, t) \exp\left[i \, \frac{2 \, q_m \, q_e}{c \, \hbar} \, \varphi\right] \tag{3.143}$$

We know that φ and $\varphi + 2\pi$ are the same point.

at φ we have:

$$\Psi_N(r, t) = \Psi_S(r, t) \exp\left[i \, \frac{2 \, q_m \, q_e}{c \, \hbar} \, \varphi\right] \tag{3.144}$$

at $\varphi + 2\pi$ we have:

$$\Psi_N(r, t) = \Psi_S(r, t) \exp\left[i \, \frac{2 \, q_m \, q_e}{c \, \hbar} \, (\varphi + 2\pi)\right] \tag{3.145}$$

To have both the two equations valid, $\frac{2 \, q_m \, q_e}{c \, \hbar}$ must be an integer n

$$\Psi_N(r, t) = \Psi_S(r, t) \exp\left[i \, n \, (\varphi + 2\pi)\right] = \Psi_S(r, t) \exp\left(i \, n \, \varphi\right) \tag{3.146}$$

This tells us that the magnetic charge is quantized:

$$q_m = \frac{c \, \hbar}{2 \, q_e} \, n \tag{3.147}$$

Q: Why charge are quantized in our universe?

A: Unknown. It is still a mystery. But, if we can find even a single magnetic monopole somewhere in our universe, the Dirac quantization condition immediately solves this puzzle for us.

$$q_e = \frac{c \, \hbar}{2 \, q_m} \, n \tag{3.148}$$

For a closed surface enclosing a magnetic monopole, no matter what gauge one uses, the vector potential must have some singularities.

If A is a non-singular function on a closed manifold, the magnetic flux through this manifold must be zero.

To proof this, we cut the manifold into two parts D_I and D_{II}

The magnetic flux though D_I is

$$\iint_{D_I} d \, S \cdot B = \iint_{D_I} d \, S \cdot \nabla \times A = \int_{\partial D_I} d \, l \cdot A \tag{3.149}$$

Here we use the Stokes' theorem and ∂D_I is the edge of D_I.

The magnetic flux though D_{II} is

$$\iint_{D_{II}} d \, S \cdot B = \iint_{D_{II}} d \, S \cdot \nabla \times A = \int_{\partial D_{II}} d \, l \cdot A \tag{3.150}$$

Therefore, the total magnetic field flux is

$$\oiint_M B \cdot d \, S = \iint_{D_I} d \, S \cdot B + \iint_{D_{II}} d \, S \cdot B = \int_{\partial D_I} d \, l \cdot A + \int_{\partial D_{II}} d \, l \cdot A \tag{3.151}$$

Notice that the edge of D_I and D_{II} are the same curve but their directions are opposite.

$$\partial D_I = -\partial D_{II} \tag{3.152}$$

$$\oiint_M B \cdot d \, S = \int_{\partial D_I} d \, l \cdot A - \int_{\partial D_I} d \, l \cdot A = 0 \tag{3.153}$$

The only way to have nonzero magnetic flux here is to have some singular vector potential. If A is singular (for all any gauge choice), we must use at least two different gauge choice to cover the whole manifold.

If we use A_I for D_I uses A_{II} for D_{II}, we get

$$\oiint_M B \cdot d S = \int_{\partial D_I} d l \cdot A_I - \int_{\partial D_I} d l \cdot A_{II} = \int_{\partial D_I} d l \cdot (A_I - A_{II}) \tag{3.154}$$

Here, I emphasize again that this singularity in A is not a physical singularity. It you measure any physical observables, there is nothing singular here. The singularity lies in A, which is NOT a physical observable. The location of this singularity changes when one choose another gauge. But no matter how one choose the gauge, there must be at least one singularity.

Later we will show that the same conclusion holds for the Berry curvature. For insulators, if the Bloch wavefunction can be defined for any k in the BZ without singularity. We can show that

$$\oiint_{BZ} \mathcal{F} \, d\vec{k} = \int_{\partial D_I} d l \cdot \mathcal{A} - \int_{\partial D_I} d l \cdot \mathcal{A} = 0. \tag{3.155}$$

Since this integral is the Hall conductivity, this means that $\sigma_{xy} = 0$ if the Bloch wave function can be defined at any k without singularity.

To have nonzero Hall conductivity, the Bloch wave functions must have some singular point in the k-spaces. As a result, we will need to define different Bloch wave functions for different regions of the k-space and these Bloch waves are connected by a gauge transformation at the boundary.

$$\left| u^{II}_{n,k} \right\rangle = e^{i \, \phi_n(k)} \left| u^I_{n,k} \right\rangle \tag{3.156}$$

We know that the Berry connection changes under this phase shift, so the Berry connections in the different regions must be different also

$$\vec{\mathcal{A}}^{II}_n = \vec{\mathcal{A}}_n^I + \nabla_{\vec{k}} \, \phi_n(k) \tag{3.157}$$

This singularity again is not a physical singularity. All physical observables show no singular behavior and their values are continuous function of k, when we go across the boundary for these different regions. The discontinuities and singularities can only be seen when you look at quantities that cannot be measured, like the phase of a wavefunction and the Berry connection. Because we use different $\mathcal{A}$ for different regions, we have

$$\oiint_{BZ} \mathcal{F} \, d\vec{k} = \int_{\partial D_I} d l \cdot (\mathcal{A}_I - \mathcal{A}_{II}) \tag{3.158}$$

which could be none zero.

The quantization of the flux is only true for a closed manifold (e.g. a sphere). For a open manifold with boundaries, the flux is in general unquantified (e.g. part of a sphere).

The same is true for the Berry curvature. Only if the integral is over the whole BZ, the total Berry curvature is quantized. In other words, the Berry curvature for insulator is quantized. But for metal, it is not.

3.9. Gauss– Bonnet theorem and topology

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.

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:

$$\oiint_M K \, d S = 2 \pi \, \chi_M \tag{3.159}$$

To define the curvature for a curve, we use a circle to fit the curve around one point on the curve. The inverse radius $\kappa=1/R$ 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 κ_1 and κ_2. The Gaussian curvature is the product of they two.

For a sphere, $\kappa_1 = \kappa_2 = 1 / R$, so $K = \frac{1}{R^2}$

For a saddle point, the surface curves up along one direction and curves down along another direction, $\kappa_1 > 0$ and $\kappa_2 < 0$. So $K = \kappa_1 \kappa_2 < 0$.

For any 2D closed manifold, the integral of $\oiint_M K \, d S$ 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 (not changing the topology), χ_M will remain the same. If one change the topology, χ_M takes *a* different value.

Sphere : $\chi_M = 2$

Torus : $\chi_M = 0$

double torus : $\chi_M = -2$

triple torus : $\chi_M = -4$

figures from wikipedia.org From left to right: sphere, torus, double torus and triple torus.

Again, it is important to emphasize here that this integral is a topological index, only if we are considering a closed manifold which has no boundary. Otherwise, it is not quantized and it is not a topological index.

3.9.1. connections to other topological properties

Topology and handles

χ_M is directly related to the genus g of the manifold.

$$\chi_M = 2 \, (1 - g). \tag{3.160}$$

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

Similarly, a sippy cup=a double torus (two handles)

a three-handled cup=a pretzel=a triple torus (three hazels)

χ_M and polyhedrons

Another definition of χ_M, if we draw a grid on the manifold,

$$\chi_M = V - E + F \tag{3.161}$$

where V, E, and F are the numbers of vertices (corners), edges and faces respectively.

For most polyhedron we are familiar with (simply connected polyhedron), they are topologically equivalent to a sphere, which has $\chi_M = 2$. So, these polyhedrons has

$$V - E + F = 2 \tag{3.162}$$

Topology and hair vortex

If we drop a (in-plane) vector at each point on this manifold, we get a vector fields, which may have vortices.

For a vortex, we can define its vorticity, which is an integer.

The total vorticity is χ_M

$$\chi_M = \sum v \tag{3.163}$$

For sphere $\chi_M = 2$, which means that total vorticity must be 2, which is nonzero!

One can't comb the hair on a 2-sphere without singularities (vortex).

If one think hair as vector fields (pointing from the end to the tip), on a sphere, this hairs must have some vortex, and the total vorticity is 2.

Examples: If one comb the hair along the longitude (or latitude) directions , there are two +1 vortex at north and south poles.

If one assumes that human hair covers the north hemisphere (of the head) and pointing downward (to -z) at the equation, which is typically try for human hairs, vorticity total is +1 (half of +2). For the majority of human beings, there is one +1 vortex. But there are more complicated cases, for example two +1 and one -1, or three +1 and two -1.

3.10. Topological index for an insulator

The topological index

$$C = \frac{1}{2\pi} \sum_n \oiint_{BZ} \mathcal{F} \, d\vec{k} \tag{3.164}$$

This integral is a topological index only if we integrate $\mathcal{F}$ over the whole BZ (a closed manifold). Because a BZ has periodic boundary conditions along x and y (for a 2D system), the BZ is a torus which is a closed manifold.

$$\mathcal{F}_n\!\left(\vec{k}\right) = \epsilon_{ij} \left\langle \partial_{ki} u_{n,k} \mid \partial_{kj} u_{n,k} \right\rangle \tag{3.165}$$

$$\sigma_{xy} = \frac{e^2}{h} \sum_{n,\text{ valence bands}} \left[\frac{1}{2\pi} \int_{BZ} d\vec{k} \, \mathcal{F}_n\!\left(\vec{k}\right) \right] + \frac{e^2}{h} \sum_{n,\text{ conduction bands}} \left[\frac{1}{2\pi} \int_{\epsilon_k < \epsilon_F} d\vec{k} \, \mathcal{F}_n\!\left(\vec{k}\right) \right] \tag{3.166}$$

For 2D systems, we will show below that the first term is quantized and is topologically invariant. The second term is not quantized and is not topologically invariant.

For insulators (first term only), the Hall conductivity is a topological index and is an Integer due to topological quantization.

For metals, the integral for the conducting band is take over only part of the BZ (the filled states, or say the Fermi sea). It is not quantized and it is not a topological index.

In other words, $\mathcal{F}$ gives us 2D topological insulators, but no topological metals.

3.11. Second quantization

3.11.1. wave functions for indistinguishable particles

Second quantization is a technique to handle indistinguishable particles.

Two distinguishable particles: particle one on state ψ_1 and particle two on state ψ_2

$$\Psi(r_1, r_2) = \psi_1(r_1)\, \psi_2(r_2) \tag{3.167}$$

n distinguishable particles:

$$\Psi(r_1, r_2, r_3, ..., r_N) = \psi_1(r_1)\, \psi_2(r_2) \, ... \psi_n(r_n) \tag{3.168}$$

Two indistinguishable particles: particle one on state ψ_1 and particle two on state ψ_2

$$\Psi(r_1, r_2) = \pm \Psi(r_2, r_1) \tag{3.169}$$

$$\Psi(r_1, r_2) = \psi_1(r_1)\, \psi_2(r_2) \pm \psi_2(r_1)\, \psi_1(r_2) \tag{3.170}$$

3 indistinguishable particles:

$$\Psi(r_1, r_2, r_3) = \psi_1(r_1)\, \psi_2(r_2)\, \psi_3(r_3) \pm \psi_1(r_1)\, \psi_3(r_3)\, \psi_2(r_2) \pm \psi_2(r_2)\, \psi_1(r_1)\, \psi_3(r_3) + \\ \psi_3(r_3)\, \psi_1(r_1)\, \psi_2(r_2) + \psi_2(r_2)\, \psi_3(r_3)\, \psi_1(r_1) \pm \psi_3(r_3)\, \psi_2(r_2)\, \psi_1(r_1) \tag{3.171}$$

n indistinguishable particles:

$$\Psi(r_1, r_2, r_3, ..., r_n) = \sum_{\mathcal{P}} (\pm 1)^{\mathcal{P}} \, \psi_{i_1}(r_1)\, \psi_{i_2}(r_2) \, ... \psi_{i_n}(r_n) \tag{3.172}$$

where $\mathcal{P}$ represents all permutations and there are n! terms here. For large n, this is an extremely complicated wavefunction. For even ten particles, n=10, there are 2.6 million terms.

Examples: If one comb the hair along the longitude (or latitude) directions , there are two +1 vortex at north and south poles.

If one assumes that human hair covers the north hemisphere (of the head) and pointing downward (to -z) at the equation, which is typically try for human hairs, vorticity total is +1 (half of +2). For the majority of human beings, there is one +1 vortex. But there are more complicated cases, for example two +1 and one -1, or three +1 and two -1.

3.10. Topological index for an insulator

The topological index

$$C = \frac{1}{2\pi} \sum_n \oiint_{BZ} \mathcal{F} \, d\vec{k} \tag{3.164}$$

This integral is a topological index only if we integrate $\mathcal{F}$ over the whole BZ (a closed manifold). Because a BZ has periodic boundary conditions along x and y (for a 2D system), the BZ is a torus which is a closed manifold.

$$\mathcal{F}_n\left(\vec{k}\right) = \epsilon_{ij} \left\langle \partial_{ki} u_{n,k} \mid \partial_{kj} u_{n,k} \right\rangle \tag{3.165}$$

$$\sigma_{xy} = \frac{e^2}{h} \sum_{n, \text{ valence bands}} \left[\frac{1}{2\pi} \int_{BZ} d\vec{k} \, \mathcal{F}_n\left(\vec{k}\right) \right] + \frac{e^2}{h} \sum_{n, \text{ conduction bands}} \left[\frac{1}{2\pi} \int_{\epsilon_k < \epsilon_F} d\vec{k} \, \mathcal{F}_n\left(\vec{k}\right) \right] \tag{3.166}$$

For 2D systems, we will show below that the first term is quantized and is topologically invariant. The second term is not quantized and is not topologically invariant.

For insulators (first term only), the Hall conductivity is a topological index and is an Integer due to topological quantization.

For metals, the integral for the conducting band is take over only part of the BZ (the filled states, or say the Fermi sea). It is not quantized and it is not a topological index.

In other words, $\mathcal{F}$ gives us 2D topological insulators, but no topological metals.

3.11. Second quantization

3.11.1. wave functions for indistinguishable particles

Second quantization is a technique to handle indistinguishable particles.

Two distinguishable particles: particle one on state ψ_1 and particle two on state ψ_2

$$\Psi(r_1, r_2) = \psi_1(r_1) \psi_2(r_2) \tag{3.167}$$

n distinguishable particles:

$$\Psi(r_1, r_2, r_3, ..., r_N) = \psi_1(r_1) \psi_2(r_2) ... \psi_n(r_n) \tag{3.168}$$

Two indistinguishable particles: particle one on state ψ_1 and particle two on state ψ_2

$$\Psi(r_1, r_2) = \pm \Psi(r_2, r_1) \tag{3.169}$$

$$\Psi(r_1, r_2) = \psi_1(r_1) \psi_2(r_2) \pm \psi_2(r_1) \psi_1(r_2) \tag{3.170}$$

3 indistinguishable particles:

$$\Psi(r_1, r_2, r_3) = \psi_1(r_1) \psi_2(r_2) \psi_3(r_3) \pm \psi_1(r_1) \psi_3(r_3) \psi_2(r_2) \pm \psi_2(r_2) \psi_1(r_1) \psi_3(r_3) + \\ \psi_3(r_3) \psi_1(r_1) \psi_2(r_2) + \psi_2(r_2) \psi_3(r_3) \psi_1(r_1) \pm \psi_3(r_3) \psi_2(r_2) \psi_1(r_1) \tag{3.171}$$

n indistinguishable particles:

$$\Psi(r_1, r_2, r_3, ..., r_n) = \sum_P (\pm 1)^P \psi_{i_1}(r_1) \psi_{i_2}(r_2) ... \psi_{i_n}(r_n) \tag{3.172}$$

where $\mathcal{P}$ represents all permutations and there are n! terms here. For large n, this is an extremely complicated wavefunction. For even ten particles, n=10, there are 2.6 million terms.

3.11.2. Fock space

The problem here is because Schrodinger equations and wavefunctions are designed for distinguishable particles and one needs to symmetries or anti-symmetries all the wavefunctions by hand.

For indistinguishable particles, it is more natural to use the particle number basis, the Fock space. In the Fock space, the many-body quantum state are written in terms of occupation numbers:

$$| \Psi \rangle = | n_1, n_2, n_3 \ldots n_N \rangle \tag{3.173}$$

where n_i is the number of particles in state $| \psi_i \rangle$. Here we don't specify which particle is in the state $| \psi_i \rangle$, only count the number of particles on this state, which makes the particle indistinguishable automatically.

3.11.3. Creation and annihilation operators:

In the Fock space, all physical operators can be written in terms of creation and annihilation operators:

For bosons

$$b_i{}^\dagger | n_1, n_2, n_3 \ldots n_N \rangle = \sqrt{n_i + 1} \; | n_1, n_2, , \ldots n_i + 1, \ldots n_N \rangle \tag{3.174}$$

$$b_i | n_1, n_2, n_3 \ldots n_N \rangle = \sqrt{n_i} \; | n_1, n_2, , \ldots n_i - 1, \ldots n_N \rangle \tag{3.175}$$

$$[b_i, b_j] = [b_i{}^\dagger, b_j{}^\dagger] = 0 \tag{3.176}$$

$$[b_i, b_j{}^\dagger] = \delta_{ij} \tag{3.177}$$

For fermions $n_i = \pm 1$ (Pauli exclusive principle)

$$c_i{}^\dagger | \ldots, 1, \ldots \rangle = 0 \tag{3.178}$$

$$c_i{}^\dagger | \ldots, 0, \ldots \rangle = | \ldots, 1, \ldots \rangle \tag{3.179}$$

$$c_i | \ldots, 1, \ldots \rangle = 0 \tag{3.180}$$

$$c_i{}^\dagger | \ldots, 0, \ldots \rangle = | \ldots, 0, \ldots \rangle \tag{3.181}$$

$$\{c_i, c_j\} = \{c_i{}^\dagger, c_j{}^\dagger\} = 0 \tag{3.182}$$

$$\{c_i, c_j{}^\dagger\} = \delta_{ij} \tag{3.183}$$

3.11.4. Particle number operator

Particle number operator for bosons $n_i = b_i{}^\dagger b_i$

$$b_i{}^\dagger b_i | \ldots, n_i, \ldots \rangle = n_i \tag{3.184}$$

Total particle number $N = \sum_i n_i = \sum_i b_i{}^\dagger b_i$

Particle number operator for fermions $n_i = c_i{}^\dagger c_i$

$$c_i{}^\dagger c_i | \ldots, 0, \ldots \rangle = 0 \tag{3.185}$$

$$c_i{}^\dagger c_i | \ldots, 1, \ldots \rangle = 1 \tag{3.186}$$

Total particle number $N = \sum_i n_i = \sum_i c_i{}^\dagger c_i$

3.11.5. Quantum states

In the Fock space, all quantum states can be written in terms of creation and annihilation operators:

Vacuum (ground states in high energy physics) $| G \rangle$. We assume that there is one and only one state in the Fock space that is annihilated by any annihilation operators. This is our vacuum state.

$$c_i \, | \, G \rangle = 0 \quad \text{for any } c_i \tag{3.187}$$

one particle states

$$\psi = \sum_i a_i \, c_i^\dagger \, | \, G \rangle \tag{3.188}$$

two particle states

$$\psi = \sum a_{ij} \, c_j^\dagger \, c_i^\dagger \, | \, G \rangle \tag{3.189}$$

n particle states

$$\psi = \sum a_{i_1 \dots i_n} \, c_{i_n}^\dagger \dots c_{i_2}^\dagger \, c_i^\dagger \, | \, G \rangle \tag{3.190}$$

3.11.6. Physical observables and expectation values

All physical observables can be written in terms of expectation values for these creation and annihilation operators:

$$X = \langle \Psi \, | \, X \, | \, \Psi \rangle = \sum f_{i1, \dots i_N} \langle G \, | \, c_{i_1} \dots c_{i_n} \, c_{i_n}^\dagger \dots c_{i_2}^\dagger \, c_i^\dagger \, | \, G \rangle \tag{3.191}$$

Remarks: these procedures are known as the "Second Quantization"

- This name are used due to historical reasons. We are not quantizing something once again. We are just using a new basis to handle indistinguishable particles.
- It is just one step away from quantum field theory. (will be discussed later)
- In both high energy and condensed matter physics, quantum field theory utilize the "Second Quantization" construction. The reason is because quantum field theory deals with more than one indistinguishable palaces, and the Second Quantization formulas are the most natural way to describe this time of physics.

3.12. tight-binding models

Fermions hopping on a lattice and each lattice site can only have n discrete quantum state. It is an simplified model for relay crystals (a discretize version of a metal). It captures most of the essential physics and provides the same physics phenomena.

One unit cell main contain multi sites (multi atoms in a unit cell), saying m sites. A site are marked by two numbers the unit cell index i=1, 2, ..., N and the site index α=1, 2, 3, ...m

Coordinate of a site: for site (i,α) the position is $x = i\,a + r_\alpha$

Bloch wave:

$$\psi_{n,k}(x) = u_{n,k}(x) \, e^{i\,k\,x} \qquad \text{continues models} \tag{3.192}$$

$$\psi_{n,\,j,\,\alpha}(k) = u_{n,\,\alpha}(k) \, \frac{e^{i\,k\,x}}{\sqrt{2\,\pi/a}} \qquad \text{tight} - \text{binding models} \tag{3.193}$$

Find the Bloch wave function = find $u_{n,\alpha}$

Berry connection and Berry curvature:

$$\mathcal{A}_{n,k} = -i \, \langle u_{n,k} \, | \, \partial_k \, | \, u_{n,k} \rangle = -i \int d\,x \, u_{n,k}{}^*(r) \, \partial_k \, u_{n,k}(r) = -i \sum_\alpha u_{n,\alpha}(k)^* \, \partial_k \, u_{n,\alpha}(k) \tag{3.194}$$

$$\mathcal{F}_{n,k} = -i \, \epsilon_{ij} \, \langle \partial_{k_i} u_{n,k} \, | \, \partial_{k_j} u_{n,k} \rangle = -i \, \epsilon_{ij} \int d\,x \, \partial_{k_i} u_{n,k}{}^*(r) \, \partial_{k_j} u_{n,k}(r) = -i \, \epsilon_{ij} \sum_\alpha \partial_{k_i} u_{n,\alpha}(k)^* \, \partial_{k_j} u_{n,\alpha}(k) \tag{3.195}$$

3.12.1. Example: a one-band model

Lets consider a simple 1D lattice with only one type of atoms.

$$H = -\sum_{ij} \left(t_{ij} \, c_i^\dagger \, c_j + t_{ij}{}^* \, c_j^\dagger \, c_i \right) + \sum_i V_i \, c_i^\dagger \, c_i \tag{3.196}$$

The first term describes the hopping from site j to i and the second term describes the hopping from i to j. Because H is Hermitian, the two coefficients must be equal to each other. Because the lattice contains only one type of atoms, V_i=constant

$$H = -\sum_{ij} \left(t_{ij} c_i^\dagger c_j + t_{ij}^* c_j^\dagger c_i \right) + V \sum_i c_i^\dagger c_i = -\sum_{ij} \left(t_{ij} c_i^\dagger c_j + t_{ij}^* c_j^\dagger c_i \right) + V \, N \tag{3.197}$$

The last term $V \, N$ is a constant. It shift the total energy by a constant, and has no other physical contribution (can be ignored if we are not interested in the total energy). In addition, we assume the electrons can only hop to the nearest neighbor: Due to translational symmetry, for all NN hoppings, the t must be the same. If we assume t is real (which is always the case for 1D NN hopping-only models)

$$H = -t \sum_{\langle ij \rangle} \left(c_i^\dagger c_j + c_j^\dagger c_i \right) = -t \sum_i c_i^\dagger c_{i+1} + h.c. \tag{3.198}$$

Fourier series:

Typically, Fourier series are applied to a periodic function in the real space, which will have a describe set of wavevectors in the k space. Here, it is the opposite. We have a continuous k-space and it is periodic (BZ), but the real space is discrete.

$$c_k = \frac{\sqrt{a}}{\sqrt{2\pi}} \sum_i c_i e^{-i k x} \tag{3.199}$$

$$c_i = \frac{\sqrt{a}}{\sqrt{2\pi}} \int_{BZ} dk \, c_k \, e^{i k x} \tag{3.200}$$

$$\{c_k, c_{k'}^\dagger\} = \left\{ \frac{1}{\sqrt{2\pi/a}} \sum_i c_i e^{-i k x_i}, \frac{1}{\sqrt{2\pi/a}} \sum_j c_j^\dagger e^{i k' x_j} \right\} =$$

$$\sum_{i,j} \{c_i, c_j^\dagger\} \frac{1}{2\pi/a} e^{-i k x_i} e^{i k' x_j} = \sum_{i,j} \delta_{i,j} \frac{1}{2\pi/a} e^{-i(k-k') x_i} = \sum_i \frac{1}{2\pi/a} e^{-i(k-k') i a} = a \, \delta((k-k') a) = \delta(k-k') \tag{3.201}$$

$$\{c_i, c_j^\dagger\} = \left\{ \frac{1}{\sqrt{2\pi/a}} \int_{BZ} dk \, c_k \, e^{i k x_i}, \frac{1}{\sqrt{2\pi/a}} \int_{BZ} dk' c_{k'}^\dagger e^{-i k' x_j} \right\} =$$

$$\frac{1}{2\pi/a} \int_{BZ} dk \int_{BZ} dk' \{c_k, c_{k'}^\dagger\} e^{i k x_i} e^{-i k' x_j} = \frac{1}{2\pi/a} \int_{BZ} dk \int_{BZ} dk' \delta(k-k') e^{i k x_i} e^{-i k' x_j} = \frac{1}{2\pi/a} \int_{BZ} dk \, e^{i k (x_i - x_j)} = \delta_{ij} \tag{3.202}$$

The Hamiltonian:

$$H = -t \sum_i c_i^\dagger c_{i+1} + h.c. \tag{3.203}$$

$$\sum_j c_j^\dagger c_{j+1} = \sum_j \frac{1}{\sqrt{2\pi/a}} \int_{BZ} dk \, c_k^\dagger e^{-i k a j} \frac{1}{\sqrt{2\pi/a}} \int_{BZ} dk' c_{k'} e^{i k'(j+1) a} =$$

$$\sum_j \frac{1}{2\pi/a} \int_{BZ} dk \int_{BZ} dk' c_k^\dagger c_{k'} e^{i k' a} e^{-i(k'-k) a j} = \int_{BZ} dk \int_{BZ} dk' c_k^\dagger c_{k'} e^{i k' a} \delta(k-k') = \int_{BZ} dk \, c_k^\dagger c_k \, e^{i k a} \tag{3.204}$$

$$H = -t \sum_i c_i^\dagger c_{i+1} + h.c. = -t \int_{BZ} dk \, c_k^\dagger c_k \, e^{i k a} + h.c. =$$

$$-t \int_{BZ} dk \, c_k^\dagger c_k \, e^{i k a} - t \int_{BZ} dk \, c_k^\dagger c_k \, e^{-i k a} = -2t \int_{BZ} dk \, c_k^\dagger c_k \cos k a = \int_{BZ} dk \, (-2t \cos k a) \, n_k \tag{3.205}$$

$$E = \int dk \, \epsilon_k \, n_k \tag{3.206}$$

Dispersion relation:

$$\epsilon_k = -2t \cos k a \tag{3.207}$$

ϵ_k is a periodic function of k. However, we only have one band.

Bloch wave:

$$\psi_n(k) = u_{n,\alpha}(k) \frac{e^{i k x}}{\sqrt{2\pi/a}} \tag{3.208}$$

$$|\psi_n(k)\rangle = \sum_x u_{n,\alpha}(k) \frac{e^{ikx}}{\sqrt{2\pi/a}} |x\rangle \tag{3.209}$$

$$c_k^\dagger |G\rangle = \sum_x u_{n,\alpha}(k) \frac{e^{ikx}}{\sqrt{2\pi/a}} c_j^\dagger |G\rangle \tag{3.210}$$

Compare with

$$c_k^\dagger = \sum_j \frac{e^{ikx_j}}{\sqrt{2\pi/a}} c_j^\dagger \tag{3.211}$$

we find that $u_k = 1$.

So we know immediately that

$$\mathcal{A} = \mathcal{F} = 0 \tag{3.212}$$

There is no Hall effect.

3.12.2. Discrete Fourier transform and Fourier series

Consider a discrete function f_i, $i = 1, 2, 3 \ldots N$

Assuming periodic boundary condition $f_{N+i} = f_i$

For large system $N \to \infty$, different boundary conditions are believe to give the same results (But there are exceptions, e.g. the edge states in topological states of matter).

$$f_j = \frac{1}{\sqrt{N}} \sum_k \tilde{f}_k \, e^{ikaj} \tag{3.213}$$

$$\tilde{f}_k = \frac{1}{\sqrt{N}} \sum_j f_j \, e^{-ikaj} \tag{3.214}$$

- Same as j, k is also a discrete variable.

This is because $f_{N+i} = f_N$

$$f_{j+N} = \frac{1}{\sqrt{N}} \sum_k \tilde{f}_k \, e^{ik(aj+aN)} = \frac{1}{\sqrt{N}} \sum_k \tilde{f}_k \, e^{ikaj} e^{ikaN} \tag{3.215}$$

$$f_j = \frac{1}{\sqrt{N}} \sum_k \tilde{f}_k \, e^{ikaj} \tag{3.216}$$

Compare the two equations, we find that $e^{ikaN} = 1$, which implies

$$k = \frac{2\pi m}{Na} = \frac{2\pi m}{L} \tag{3.217}$$

For infinite systems, $L \to \infty$, the discrete sum turns into an integral.

- Same as j, k also have a periodicity, and the periodicity is also N.

$$\tilde{f}_{k_m} e^{ik_m a j} = \tilde{f}_{k_m} e^{i\frac{2\pi}{Na}m a j} = \tilde{f}_{k_m} e^{i\frac{2\pi}{N}m j} \tag{3.218}$$

$$\tilde{f}_{k_{m+N}} e^{ik_{m+N} a j} = \tilde{f}_{k_{m+N}} e^{i\frac{2\pi}{Na}(m+N)a j} = \tilde{f}_{k_{m+N}} e^{i\frac{2\pi}{Na}m a j} e^{i\frac{2\pi}{Na}N a j} = \tilde{f}_{k_{m+N}} e^{i\frac{2\pi}{N}m j} e^{i2\pi j} = \tilde{f}_{k_{m+N}} e^{i\frac{2\pi}{N}m j} \tag{3.219}$$

$$\tilde{f}_{k_m} = \tilde{f}_{k_{m+N}} \tag{3.220}$$

$$m = -\frac{N}{2}, -\frac{N}{2}+1, \ldots \frac{N}{2}-1, \tag{3.221}$$

$$k = -\frac{\pi}{L}, \ -\frac{\pi}{L} + 1 \times \frac{2\pi}{L}, \ -\frac{\pi}{L} + 2 \times \frac{2\pi}{L}, \ \dots \frac{\pi}{L} - \frac{2\pi}{L},$$ (3.222)

- Identities:

$$\frac{1}{N} \sum_k e^{ikaj} \, e^{-ikaj'} = \delta_{j,j'}$$ (3.223)

$$\frac{1}{N} \sum_j e^{ikaj} \, e^{-ik'aj} = \delta_{k,k'}$$ (3.224)

3.12.3. Example: a two-band model in 1D

$$H = -t \sum_i \left(a_i^\dagger b_i + b_i^\dagger a_{i+1} + h.c. \right) + V_a \sum_i a_i^\dagger a_i + V_b \sum_i b_i^\dagger b_i$$ (3.225)

Position of a sites in the jth unit cells:

$$r = a \times j + r_a$$ (3.226)

Position of b sites in the jth unit cells:

$$r = b \times j + r_b$$ (3.227)

$$a_k = \frac{1}{\sqrt{N}} \sum_i a_i \, e^{ikx}$$ (3.228)

$$a_i = \frac{1}{\sqrt{N}} \sum_k a_k \, e^{-ikx}$$ (3.229)

$$b_k = \frac{1}{\sqrt{N}} \sum_i a_i \, e^{ikx}$$ (3.230)

$$b_i = \frac{1}{\sqrt{N}} \sum_k a_k \, e^{-ikx}$$ (3.231)

$$\sum_j a_j^\dagger b_j = \sum_j \frac{1}{\sqrt{N}} \sum_k a_k^\dagger \, e^{-ik(aj+r_a)} \frac{1}{\sqrt{N}} \sum_{k'} b_{k'} \, e^{ik'(aj+r_b)} = \frac{1}{N} \sum_k \sum_{k'} a_k^\dagger \, b_{k'} \sum_j e^{-ik(aj+r_a)} \, e^{ik'(aj+r_b)} =$$

$$\sum_k \sum_{k'} a_k^\dagger \, b_{k'} \frac{1}{N} \sum_j e^{-i(k-k')aj} \, e^{-ikr_a} \, e^{ik'r_b} = \sum_k \sum_{k'} a_k^\dagger \, b_{k'} \, \delta_{k,k'} \, e^{-ikr_a} \, e^{ik'r_b} = \sum_k a_k^\dagger \, b_k \, e^{ik(r_b-r_a)}$$ (3.232)

$$\sum_j b_j^\dagger \, a_{j+1} = \sum_k a_k^\dagger \, b_k \, e^{ik(r_a-r_b)}$$ (3.233)

Assume that $r_b - r_a = \frac{a}{2}$

$$\sum_j a_j^\dagger \, b_j = \sum_k a_k^\dagger \, b_k \, e^{ika/2}$$ (3.234)

$$\sum_j b_j^\dagger \, a_{j+1} = \sum_k b_k^\dagger \, a_k \, e^{ika/2}$$ (3.235)

We also know that

$$\sum_i a_i^\dagger \, a_i = \sum_k a_k^\dagger \, a_k$$ (3.236)

Therefore, in the k-space, the Hamiltonian looks like

$$H = -t \sum_k \left(a_k^\dagger \, b_k \, e^{ika/2} + b_k^\dagger \, a_k \, e^{ika/2} + b_k^\dagger \, a_k \, e^{-ika/2} + a_k^\dagger \, b_k \, e^{-ika/2} \right) + V_a \sum_k a_k^\dagger \, a_k + V_b \sum_k b_k^\dagger \, b_k =$$

$$-2t \sum_k \left[\left(a_k^\dagger \, b_k \cos\left(\frac{ka}{2}\right) + b_k^\dagger \, a_k \cos\left(\frac{ka}{2}\right) \right) + V_a \, a_k^\dagger \, a_k + V_b \, b_k^\dagger \, b_k \right]$$ (3.237)

We can write it in a matrix form:

$$H = \sum_k \begin{pmatrix} a_k^\dagger, & b_k^\dagger \end{pmatrix} \begin{pmatrix} V_a & -2\,t\,\cos\!\left(\frac{k\,a}{2}\right) \\ -2\,t\,\cos\!\left(\frac{k\,a}{2}\right) & V_b \end{pmatrix} \begin{pmatrix} a_k \\ b_k \end{pmatrix} \tag{3.238}$$

- For any tight-binding models with two quantum states per unit cell, the Hamiltonian can be written in terns of a two-by-two matrix in the k space like this:

$$H = \sum_k \begin{pmatrix} a_k^\dagger, & b_k^\dagger \end{pmatrix} \mathcal{H}(k) \begin{pmatrix} a_k \\ b_k \end{pmatrix} \tag{3.239}$$

where $\mathcal{H}$ is a 2×2 Hermitian matrix as a function of k. It is called the kernel of the Hamiltonian. It contains all the useful information of a Hamiltonian but it is NOT the Hamiltonian, although in literature, it is usually called a Hamiltonian.

For the model considered here,

$$\mathcal{H}(k) = \begin{pmatrix} V_a & -2\,t\,\cos\!\left(\frac{k\,a}{2}\right) \\ -2\,t\,\cos\!\left(\frac{k\,a}{2}\right) & V_b \end{pmatrix} \tag{3.240}$$

For more generic cases, if one have m quantum state per unit cell, this $\mathcal{H}$(k) will become a m×m Hermitian matrix. It captures all the information about the system, including the topological nature.

The bottom line: for any tight-binding models, at the end of the day, one will get a n×n matrix $\mathcal{H}$(k) which contains all the information of the Hamiltonian.

When we have a matrix, we know what to do to get eigenvalue and eigenfunctions of the Hamiltonian: we diagonalize the matrix.

At each k point, one can define a unitary transformation

$$\begin{pmatrix} c_k \\ d_k \end{pmatrix} = U_k^{-1} \begin{pmatrix} a_k \\ b_k \end{pmatrix} \tag{3.241}$$

$$\begin{pmatrix} c_k^\dagger & d_k^\dagger \end{pmatrix} = \begin{pmatrix} a_k^\dagger & b_k^\dagger \end{pmatrix} U_k \tag{3.242}$$

where $U_k^{-1} = U_k^\dagger$

$$H = \sum_k \begin{pmatrix} c_k^\dagger, & d_k^\dagger \end{pmatrix} U_k^+ \, \mathcal{H} \, U_k \begin{pmatrix} c_k \\ d_k \end{pmatrix} \tag{3.243}$$

If we choose U_k such that $U_k^+ \, \mathcal{H} \, U_k$ is a diagonal matrix,

$$U_k \, \mathcal{H} \, U_k^{-1} = \begin{pmatrix} \epsilon_c & 0 \\ 0 & \epsilon_d \end{pmatrix} \tag{3.244}$$

(assume $\epsilon_c < \epsilon_d$)

$$H = \sum_k \begin{pmatrix} c_k^\dagger, & d_k^\dagger \end{pmatrix} \begin{pmatrix} \epsilon_c & 0 \\ 0 & \epsilon_d \end{pmatrix} \begin{pmatrix} c_k \\ d_k \end{pmatrix} = \sum_k \epsilon_c(k)\, c_k^\dagger\, c_k + \sum_k \epsilon_d(k)\, d_k^\dagger\, d_k \tag{3.245}$$

We know that using Bloch waves (which are eigenstates of the Hamiltonian), the total energy is

$$E = \sum_{k,n} \epsilon_n(k)\, n_n(k) = \sum_k \epsilon_1\, n_1(k) + \sum_k \epsilon_2\, n_2(k) + \sum_k \epsilon_3\, n_3(k) + \dots \tag{3.246}$$

Therefore, in terms of Bloch waves, the Hamiltonian should take the form:

$$H = \sum_k \epsilon_1(k)\, \gamma_{1,k}^\dagger\, \gamma_{1,k} + \sum_k \epsilon_2(k)\, \gamma_{2,k}^\dagger\, \gamma_{2,k} + \sum_k \epsilon_3(k)\, \gamma_{3,k}^\dagger\, \gamma_{3,k} + \dots \tag{3.247}$$

where $\gamma_{n,k}^\dagger$ is the creation operator for a bloch wave in band n at momentum k.

Compare with Eq. 245, we find that in this model we get a two energy bands (c and d). The dispersions are $\epsilon_c(k)$ and $\epsilon_d(k)$ respectively. $c^\dagger$ creates a Bloch wave with momentum k in the lower band. $d_k^\dagger$ creates a Bloch wave with momentum k in the upper bands.

For the problem we considered here, the dispersions are

$$\epsilon_c = \frac{V_a + V_b}{2} - \sqrt{\left[2\,t\,\cos\!\left(\frac{k\,a}{2}\right)\right]^2 + \left(\frac{V_a - V_b}{2}\right)^2} \tag{3.248}$$

$$\epsilon_d = \frac{V_a + V_b}{2} + \sqrt{\left[2\,t\cos\left(\frac{k\,a}{2}\right)\right]^2 + \left(\frac{V_a - V_b}{2}\right)^2} \tag{3.249}$$

Q: What determines the number of bands n?

A: Number of allowed quantum states in a unit cells.

In general, if each unit cell contains m quantum states described by the fermion operators $c_1, c_2 \dots c_m$, the tight-binding model will contains m energy bands and the kernel $\mathcal{H}_k$ will be a m×m matrix.

Q: What can we learn from the eigenvectors of $\mathcal{H}(k)$.

A: $u_{n,k}$, the coeffiicent for the Bloch waves.

We know that a Bloch wave in a continous space is:

$$\psi_{n,k}(x) = u_{n,k}(x)\,\frac{e^{i\,k\,x}}{\sqrt{N}} \tag{3.250}$$

For tight-binding models, $u_{n,k}(x)$ is now a discrete function in the real space. Due to the periodic structure, we only need to know the value $u_{n,k}(x)$ in one unit cell. In other works, we just need to know the value of this function at site a and b for the two band model we considered here (only two values). Therefore, $u_{n,k}$ is now a two component vector:

$$u_{n,k}(x) = (u_{i,a}(k),\, u_{i,b}(k)) \tag{3.251}$$

The eignevectors of $\mathcal{H}(k)$ are also a 2-component vector:

$$\mathcal{H}\begin{pmatrix} v^{(-)}{}_a(k) \\ v^{(-)}{}_b(k) \end{pmatrix} = \epsilon_c \begin{pmatrix} v^{(-)}{}_a(k) \\ v^{(-)}{}_b(k) \end{pmatrix} \tag{3.252}$$

$$\mathcal{H}\begin{pmatrix} v^{(+)}{}_a(k) \\ v^{(+)}{}_b(k) \end{pmatrix} = \epsilon_d \begin{pmatrix} v^{(+)}{}_a(k) \\ v^{(+)}{}_b(k) \end{pmatrix} \tag{3.253}$$

And in fact, these eigenvectors are percisely the coefficent of the Bloch waves ($u_{n,k}$).

$$c_k{}^\dagger = v^{(-)}{}_a(k)\,a_k{}^\dagger + v^{(-)}{}_b(k)\,b_k{}^\dagger \tag{3.254}$$

$$d_k{}^\dagger = v^{(+)}{}_a(k)\,a_k{}^\dagger + v^{(+)}{}_b(k)\,b_k{}^\dagger \tag{3.255}$$

For more generic cases, for a m-band model, there are m eigenvectors $u_{n,\alpha}(k)$ with n=1,2,…m, where n is the band index. For each band n, the eignevector is precisely the coeffiecient $u_{n,k}$ of the Bloch wave for this band.

In the continoum we have $u_{n,k}(x)$. For tight-binding models, we have $u_n(k) = (u_{n,1}(k),\, u_{n,2}(k) \dots u_{n,m}(k))$.

For each bands, we can define $\mathcal{A}$

For the continous case, we have

$$\mathcal{A}_n = -i\,\langle u_n(k)\,|\,\partial_k\,|\,u_n(k)\rangle = -i\int d\!r\, u_n(k,\,r)^*\,\partial_k\,u_n(k,\,r) \tag{3.256}$$

For tight-binding models, we have

$$\mathcal{A}_n = -i\,\langle u_n(k)\,|\,\partial_k\,|\,u_n(k)\rangle = -i\sum_\alpha u_{n,\alpha}(k)^*\,\partial_k\,u_{n,\alpha}(k) \tag{3.257}$$

3.12.4. Summary

- Tight-Binding model: a discrete version of Bloch waves
- The key properties of a tight-binding model is coded in $\mathcal{H}_k$, which is a m×m Hermitian matrix. Each component is a function of k. (In other words, $\mathcal{H}_k$ is a matrix function of k.)
- The eigenvalues if $\mathcal{H}$ (as a function of k) gives the band structure (the dispersion relation) for each bands $\epsilon_n(k)$ with $n = 1, 2, \dots m$
- The eigenvectors as a function of k gives the Bloch wave: $u_{n,\alpha}(k)$, where n=1,2,…, m is the band index and α=1,2…m marks different site in a unit cell.

- Using these $u_{n,\alpha}(k)$, we can define the Berry connection, Berry curvature and the Chern number. Just replace the integral in the real space by summing over i=1...n

Before 80s, people will just stop at step #3, without calculating the eigenvectors. This is because eigenvectors are the wavefunction, which cannot be measured directly in experiments. So people only care about the eigenvalues, which is the dispersion relation (a physical observable).

After the discovery of topological insulators, physicists realized that although the wavefunction can not be measured directly, it contains the topological information, which is a physical observable (the Hall conductivity).

3.13. Example: a topologically-nontrival model

Example:

$$\mathcal{H} = \begin{pmatrix} -2\,t\cos k_x - 2\,t\cos k_y - \mu & \Delta\left(\sin k_x - i\,\sin k_y\right) \\ \Delta\left(\sin k_x + i\,\sin k_y\right) & 2\,t\cos k_x + 2\,t\cos k_y + \mu \end{pmatrix} \tag{3.258}$$

Here, we assume $\Delta > 0$ and $t > 0$ and $-4\,|t| < \mu < 4\,|t|$.

This is the Hamiltonian of a p+i p superconductor, which will be discussed later. Although it is not the Hamiltonian of an insulator, the physics are essentially the same.

Here, I will not discuss about where this Hamiltonian comes from. We just use it as an example to demonstrate the nontrivial topological index. Let's assume that in some tight-binding model, we get a kernal of the Hamiltonian like this.

Here I use μ to refer to some control parameter, which is NOT the chemical potential.

3.13.1. More general case

For any two-band models, $\mathcal{H}$ is a two-by-two Hermitian matrix.

For a two-by-two Hermitian matrix, one can always separate it into I, and σ_i

$$\mathcal{H} = \mathcal{H}_0(k)\,I + \mathcal{H}_x(k)\,\sigma_x + \mathcal{H}_y(k)\,\sigma_y + \mathcal{H}_z(k)\,\sigma_z \tag{3.259}$$

$\mathcal{H}_0(k)$, $\mathcal{H}_x(k)$, $\mathcal{H}_y(k)$ and $\mathcal{H}_z(k)$ are real functions of k. If we define $\vec{\mathcal{H}}(k) = \left(\mathcal{H}_x(k),\ \mathcal{H}_y(k),\ \mathcal{H}_z(k)\right)$,

$$\mathcal{H} = \mathcal{H}_0(k)\,I + \vec{\mathcal{H}}(k)\cdot\vec{\sigma} \tag{3.260}$$

This is rather similar to a spin S=1/2 in a magnetic field.

$$H = \text{constant} + \mu\,\vec{B}\cdot\vec{\sigma} \tag{3.261}$$

For the case studied here,

$$\mathcal{H}_0(k) = 0 \tag{3.262}$$

$$\mathcal{H}_x(k) = \Delta \sin\left(k_x\right) \tag{3.263}$$

$$\mathcal{H}_y(k) = \Delta \sin\left(k_y\right) \tag{3.264}$$

$$\mathcal{H}_z(k) = -2\,t\cos k_x - 2\,t\cos k_y - \mu \tag{3.265}$$

The eigenvalues of $\mathcal{H}$ are

$$E_\pm = \mathcal{H}_0(k) \pm \left|\vec{\mathcal{H}}(k)\right| = \mathcal{H}_0(k) \pm \sqrt{\mathcal{H}_x(k)^2 + \mathcal{H}_y(k)^2 + \mathcal{H}_z(k)^2} \tag{3.266}$$

Pretty much the same as the spin case, where

$$E = \text{constant} + \mu\left|\vec{B}\right| \tag{3.267}$$

For the top band, $E_+ \geq \mathcal{H}_0$, for the lower band $E_- \leq \mathcal{H}_0$. The energy gap is:

$$\Delta(k) = E_+(k) - E_-(k) = 2\left|\vec{\mathcal{H}}(k)\right| \tag{3.268}$$

As long as $\vec{\mathcal{H}}(k) \neq 0$, the two band will not cross with each other.

Let's focus on the lower band (E_-), its eigenvector is

$$u^{(I)}{}_-(k) = \frac{1}{\mathcal{N}^{(I)}} \begin{pmatrix} \mathcal{H}_z(k) - \left| \vec{\mathcal{H}}(k) \right| \\ \mathcal{H}_x(k) + i\,\mathcal{H}_y(k) \end{pmatrix} \tag{3.269}$$

This wavefunction is singular if $\mathcal{H}_x(k) = \mathcal{H}_y(k) = 0$ and $\mathcal{H}_z > 0$,

$$u^{(I)}{}_-(k) = \frac{1}{\mathcal{N}^{(I)}} \begin{pmatrix} \mathcal{H}_z - \left| \vec{\mathcal{H}}(k) \right| \\ \mathcal{H}_x(k) + i\,\mathcal{H}_y(k) \end{pmatrix} = \frac{1}{\mathcal{N}^{(I)}} \begin{pmatrix} \mathcal{H}_z - \sqrt{\mathcal{H}_z(k)^2} \\ 0 \end{pmatrix} = \frac{1}{\mathcal{N}^{(I)}} \begin{pmatrix} \mathcal{H}_z - |\mathcal{H}_z(k)| \\ 0 \end{pmatrix} = \frac{1}{\mathcal{N}^{(I)}} \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{3.270}$$

This is indeed a problem for the Hamiltonian shown above. At $k_x = k_y = \pi$.

$$\mathcal{H}_0(k) = 0 \tag{3.271}$$

$$\mathcal{H}_x(k) = \Delta \sin(k_x) = 0 \tag{3.272}$$

$$\mathcal{H}_y(k) = \Delta \sin(k_y) = 0 \tag{3.273}$$

$$\mathcal{H}_z(k) = -2\,t \cos k_x - 2\,t \cos k_y - \mu = 4\,t - \mu > 0 \tag{3.274}$$

In fact, there is another way to write down the same eigenvector (a phase shift)

$$u^{(II)}{}_-(k) = \frac{1}{\mathcal{N}^{(I)}} \begin{pmatrix} \mathcal{H}_z - \left| \vec{\mathcal{H}}(k) \right| \\ \mathcal{H}_x(k) + i\,\mathcal{H}_y(k) \end{pmatrix} \times \frac{\frac{\mathcal{H}_z + \left| \vec{\mathcal{H}}(k) \right|}{\mathcal{H}_x(k) + i\,\mathcal{H}_y(k)}}{\left| \frac{\mathcal{H}_z + \left| \vec{\mathcal{H}}(k) \right|}{\mathcal{H}_x(k) + i\,\mathcal{H}_y(k)} \right|} = \frac{1}{\mathcal{N}^{(II)}} \begin{pmatrix} \frac{-\mathcal{H}_x(k)^2 - \mathcal{H}_y(k)^2}{\mathcal{H}_x(k) + i\,\mathcal{H}_y(k)} \\ \mathcal{H}_z + \left| \vec{\mathcal{H}}(k) \right| \end{pmatrix} = \frac{1}{\mathcal{N}^{(II)}} \begin{pmatrix} -\mathcal{H}_x(k) + i\,\mathcal{H}_y(k) \\ \mathcal{H}_z + \left| \vec{\mathcal{H}}(k) \right| \end{pmatrix} \tag{3.275}$$

$$u^{(II)}{}_-(k) = u^{(I)}{}_-(k)\, e^{i\,\phi(k)} \tag{3.276}$$

This new wavefunction is well defined at $\mathcal{H}_x(k) = \mathcal{H}_y(k) = 0$ and $\mathcal{H}_z(k) > 0$. However, it is NOT well defined at $\mathcal{H}_x(k) = \mathcal{H}_y(k) = 0$ and $\mathcal{H}_z(k) < 0$,

$$u^{(II)}{}_-(k) = \frac{1}{\mathcal{N}^{(II)}} \begin{pmatrix} -\mathcal{H}_x(k) + i\,\mathcal{H}_y(k) \\ \mathcal{H}_z(k) + \left| \vec{\mathcal{H}}(k) \right| \end{pmatrix} = \frac{1}{\mathcal{N}^{(II)}} \begin{pmatrix} 0 \\ \mathcal{H}_z(k) + \sqrt{\mathcal{H}_z(k)^2} \end{pmatrix} = \frac{1}{\mathcal{N}^{(II)}} \begin{pmatrix} 0 \\ \mathcal{H}_z(k) + |\mathcal{H}_z(k)| \end{pmatrix} = \frac{1}{\mathcal{N}^{(II)}} \begin{pmatrix} 0 \\ \mathcal{H}_z(k) - \mathcal{H}_z(k) \end{pmatrix} = 0 \tag{3.277}$$

For the Hamiltonian shown above, the origin is such a point. $k_x = k_y = 0$

$$\mathcal{H}_0(k) = 0 \tag{3.278}$$

$$\mathcal{H}_x(k) = \Delta \sin(k_x) = 0 \tag{3.279}$$

$$\mathcal{H}_y(k) = \Delta \sin(k_y) = 0 \tag{3.280}$$

$$\mathcal{H}_z(k) = -2\,t \cos k_x - 2\,t \cos k_y - \mu = -4\,t - \mu < 0 \tag{3.281}$$

Therefore, we need to cut the BZ into two areas and use two different wave functions to describe the Bloch waves. They are connected by a gauge transformation

$$u^{(II)}{}_-(k) = u^{(I)}{}_-(k)\, e^{i\,\phi(k)} \tag{3.282}$$

$$\mathcal{A}_-{}^{(II)}(k) = \mathcal{A}_-{}^{(I)}(k) + \nabla_k \phi(k) \tag{3.283}$$

This is in strong analogy to the magnetic monopole case.

Some comments:

- These singularities are NOT physical. If one measure any physical observables, there is no singularity anywhere in the momentum space. However. for the wavefunction and the Berry connection, which are not physical observables, there is always some singularity for this Hamiltonian.
- The location of these singularity points depends on the gauge (phase) choice. In other words, the location of the singularities has no physical meaning either.

■ Only one thing about these singularities are physical. That is: there must be some singularity points. This statement is independent of gauge choice and it tells us that the topological index is nonzero.

Let's compute the topological index for this model. For these model, there are four special points which satisfy $\mathcal{H}_x = \mathcal{H}_y = 0$. They are $\vec{k} = (0, 0), \vec{k} = (\pi, \pi), \vec{k} = (0, \pi)$ and $\vec{k} = (\pi, 0)$

At these four points, the values of $\mathcal{H}_z$ are: $\mathcal{H}_z = -4\,t - \mu,\, 4\,t - \mu,\, -\mu$ and $-\mu$ respectively.

3.13.2. case I: $\mu < -4\,t$

If $\mu < -4\,t$, all the four special points has $\mathcal{H}_z > 0$. So we can use $u^{(II)}_-(k)$ for the whole BZ, and there is no singularity points.

$$u^{(II)}_-(k) = \frac{1}{\mathcal{N}^{(I)}} \begin{pmatrix} \mathcal{H}_z(k) - \left| \vec{\mathcal{H}}(k) \right| \\ \mathcal{H}_x(k) + i\,\mathcal{H}_y(k) \end{pmatrix} \tag{3.284}$$

Then we can get Berry connection

$$\mathcal{A}_-^{(II)} = -i\, \langle u_-(k) \mid \partial_k \mid u_-(k) \rangle = -i\, \left[u^{(II)}_-(k) \right]^\dagger \partial_k\, u^{(II)}_-(k) = -i\, \sum_{\alpha=1}^{2} u^{(II)}_{-,\alpha}(k)^* \,\partial_k\, u_{-,\alpha}(k) \tag{3.285}$$

where $\left[u^{(I)}_-(k) \right]^\dagger$ is the Hermitian conjugate of $u^{(I)}_-(k)$, i.e. complex conjugate and transpose. $\left[u^{(I)}_-(k) \right]^\dagger$ is a row vector, while $u^{(I)}_-(k)$ is a column vector.

The Berry curvature:

$$\Omega_- = \nabla \times \mathcal{A}_- \tag{3.286}$$

The total Berry curvature (2π times the Chern number)

$$\oiint_{BZ} d\,k\, \Omega_- = \oiint_{BZ} d\,k\, \nabla \times \mathcal{A}_- = \oint_{\partial BZ} d\,k\, \mathcal{A}_- = 0 \tag{3.287}$$

The Chern number:

$$C = \frac{1}{2\,\pi} \oiint_{BZ} d\,k\, \Omega_- = 0. \tag{3.288}$$

3.13.3. Marginal case: $\mu = -4\,t$

$\mathcal{H}_x(k) = \mathcal{H}_y(k) = \mathcal{H}_z(k) = 0$ at k=(π,π)

Because the gap between the two bands is

$$\Delta(k) = E_+ - E_- = 2\left| \vec{\mathcal{H}}(k) \right| \tag{3.289}$$

$\Delta(k) = 0$ at k=(π,π).

The gap closes at k=(π,π), i.e. the two energy bands touch each other at k=(π,π). This gives us a Dirac point.

Not an insulator (no gap). So we cannot define a topological index.

3.13.4. case II: $-4\,t < \mu < 0$

For $-4\,t < \mu < 0$, $\mathcal{H}_z < 0$ at k=(0,0), and $\mathcal{H}_z > 0$ for all three other points.

Therefore, we need two wave-functions.

Near k=(0,0), we use

$$u^{(I)}_-(k) = \frac{1}{\mathcal{N}^{(I)}} \begin{pmatrix} \mathcal{H}_z(k) - \left| \vec{\mathcal{H}}(k) \right| \\ \mathcal{H}_x(k) + i\,\mathcal{H}_y(k) \end{pmatrix} \tag{3.290}$$

Near the other three special points, we use

$$u^{(\text{II})}{}_-(k) = \frac{1}{\mathcal{N}^{(I)}} \begin{pmatrix} \mathcal{H}_z(k) - \left| \vec{\mathcal{H}}(k) \right| \\ \mathcal{H}_x(k) + i\,\mathcal{H}_y(k) \end{pmatrix} \tag{3.291}$$

We can cut the BZ into two parts using a circle around the origin. In side the circle, D_I, we use $u^{(I)}$. Outside the circle we use $u^{(\text{II})}$. This will give as two different Berry connections. In region D^I, we have $A^{(I)}$, and in region D_{II} we have $A^{(\text{II})}$. The are connected by a gauge transformation:

$$u^{(\text{II})}{}_-(k) = u^{(I)}{}_-(k)\, e^{i\,\phi(k)} \tag{3.292}$$

$$\mathcal{A}_-^{(\text{II})}(k) = \mathcal{A}_-^{(I)}(k) + \nabla_k\,\phi(k) \tag{3.293}$$

The total Berry curvature (2π times the Chern number)

$$\oiint_{\text{BZ}} d\,k\,\Omega_- = \int\!\!\int_{D_I} d\,k\,\nabla\times\mathcal{A}_-^{(I)} + \int\!\!\int_{D_{\text{II}}} d\,k\,\nabla\times\mathcal{A}_-^{(\text{II})} = \oint_{\partial D_I} d\,k\,\mathcal{A}_-^{(I)} + \oint_{\partial D_{\text{II}}} d\,k\,\mathcal{A}_-^{(\text{II})} = \oint_{\partial D_I} d\,k\,\mathcal{A}_-^{(I)} - \oint_{\partial D_I} d\,k\,\mathcal{A}_-^{(\text{II})} =$$

$$\oint_{\partial D_I} d\,k\big(\mathcal{A}_-^{(I)} - \mathcal{A}_-^{(\text{II})}\big) = -\oint_{\partial D_I} d\,k\,\nabla_k\,\phi(k) = -\int_0^{2\pi} d\,\theta\,\partial_\theta\,\phi = -[\phi(\theta = 2\,\pi) - \phi(\theta = 0)] = \phi(\theta = 0) - \phi(\theta = 2\,\pi) \tag{3.294}$$

$$\phi(\theta = 0) - \phi(\theta = 2\,\pi) = 2\,\pi\,n \tag{3.295}$$

Therefore, the Chern number is quantized.

For the model we considered here:

$$\mathcal{H}_0(k) = 0 \tag{3.296}$$

$$\mathcal{H}_x(k) = \Delta \sin(k_x) \tag{3.297}$$

$$\mathcal{H}_y(k) = \Delta \sin(k_y) \tag{3.298}$$

$$\mathcal{H}_z(k) = -2\,t\cos k_x - 2\,t\cos k_y - \mu \tag{3.299}$$

If we choose the boundary between D_I and D_{II} to be a very small circle ($k \sim 0$)

$$\mathcal{H}_x(k) = \Delta \sin(k_x) \approx \Delta\,k_x + O\big(k^2\big) \tag{3.300}$$

$$\mathcal{H}_y(k) = \Delta \sin(k_y) \approx \Delta\,k_y + O\big(k^2\big) \tag{3.301}$$

$$\mathcal{H}_z(k) = -2\,t\cos k_x - 2\,t\cos k_y - \mu = -2\,t - \mu + O\big(k^2\big) \tag{3.302}$$

$$e^{i\,\phi(k)} = \frac{\frac{\mathcal{H}_z + \left|\vec{\mathcal{H}}(k)\right|}{\mathcal{H}_x(k) + i\,\mathcal{H}_y(k)}}{\left|\frac{\mathcal{H}_z + \left|\vec{\mathcal{H}}(k)\right|}{\mathcal{H}_x(k) + i\,\mathcal{H}_y(k)}\right|} = \frac{\frac{1}{\mathcal{H}_x(k) + i\,\mathcal{H}_y(k)}}{\left|\frac{1}{\mathcal{H}_x(k) + i\,\mathcal{H}_y(k)}\right|} = \frac{\left|\mathcal{H}_x(k) + i\,\mathcal{H}_y(k)\right|}{\mathcal{H}_x(k) + i\,\mathcal{H}_y(k)} = \frac{\left|\Delta\,k_x + i\,\Delta\,k_y\right|}{\Delta\,k_x + i\,\Delta\,k_y} = \frac{\left|k_x + i\,k_y\right|}{k_x + i\,k_y} = \frac{\left|k\,\|\,e^{i\,\theta}\right|}{\left|k\right|\,e^{i\,\theta}} = \frac{1}{e^{i\,\theta}} = e^{-i\,\theta} \tag{3.303}$$

$$\phi(k) = -\theta \tag{3.304}$$

$$\oiint_{\text{BZ}} d\,k\,\Omega_- = \phi(\theta = 0) - \phi(\theta = 2\,\pi) = 0 - (-2\,\pi) = 2\,\pi \tag{3.305}$$

$$C = \frac{1}{2\,\pi} \oiint_{\text{BZ}} d\,k\,\Omega_- = 1 \tag{3.306}$$

3.13.5. Marginal case: $\mu = 0$

Two Dirac points at k=$(0,\pi)$ and $(\pi,0)$. Not an insulator. No topological index.

3.13.6. case III: $0 < \mu < 4\,t$

For $0 < \mu < 4\,t$, $\mathcal{H}_z > 0$ at k=(π,π), and $\mathcal{H}_z < 0$ for all three other points.

Use a small circle centered at (π,π) to cut the system into two parts. Inside the circle, near (π,π), we use u^{II} and outside the circle, near $(0,0)$ we use u^I.

Chern number C=1.

3.13.7. Marginal case: $\mu = 4\,t$

Dirac points at k=(π,π). Not an insulator. No topological index.

3.13.8. case IV: $\mu > 4\,t$

All the four special points has $\mathcal{H}_z < 0$. We just use u^I for the whole BZ. No singularity. C=0

3.13.9. case IV: $\mu > 4\,t$

All the four special points has $\mathcal{H}_z < 0$. We just use u^I for the whole BZ. No singularity. C=0

3.13.10. the top band?

The top band has the opposite Chern number $C_+ = -C_-$.

In fact, one can prove that for any tight-binding models, the total Chern number for all the bands is always 0.

So here, we have $C_+ + C_- = 0$.

3.13.11. Summary

- This model has four phases. Two topological phases with C=+1 and two trivial insulator phase with C=0.
- A topological phase and the trivial insulator phase are separated by a phase transition point, which is known as a topological phase transition.
  - Across the topological phase transition, the topological index changes its value
  - Across a topological transition, the insulating gap closes and then reopens (generically true)
  - This is actually a generic statement: A band can only change its Chern number by crossing with another band. i.e. At the topological transition point, the system must be a metal or a semi-metal (with no insulating gap).
- Gap closing is a necessary condition for a topological transition, but it is not a sufficient one. One may close and reopen the gap without changing the topological index. e.g. the μ=0 point here.

3.14. Some properties of the Berry curvature $\mathcal{F}$ (we will limit our discussions to 2D systems)

3.13.7. Marginal case: $\mu = 4\,t$

Dirac points at k=(π,π). Not an insulator. No topological index.

3.13.8. case IV: $\mu > 4\,t$

All the four special points has $\mathcal{H}_z < 0$. We just use u^I for the whole BZ. No singularity. C=0

3.13.9. case IV: $\mu > 4\,t$

All the four special points has $\mathcal{H}_z < 0$. We just use u^I for the whole BZ. No singularity. C=0

3.13.10. the top band?

The top band has the opposite Chern number $C_+ = -C_-$.

In fact, one can prove that for any tight-binding models, the total Chern number for all the bands is always 0.

So here, we have $C_+ + C_- = 0$.

3.13.11. Summary

- This model has four phases. Two topological phases with C=+1 and two trivial insulator phase with C=0.
- A topological phase and the trivial insulator phase are separated by a phase transition point, which is known as a topological phase transition.
  - Across the topological phase transition, the topological index changes its value
  - Across a topological transition, the insulating gap closes and then reopens (generically true)
  - This is actually a generic statement: A band can only change its Chern number by crossing with another band. i.e. At the topological transition point, the system must be a metal or a semi-metal (with no insulating gap).
- Gap closing is a necessary condition for a topological transition, but it is not a sufficient one. One may close and reopen the gap without changing the topological index. e.g. the μ=0 point here.

3.14. The model of Haldane on a honeycomb lattice

3.14.1. Honeycomb lattice (graphene)
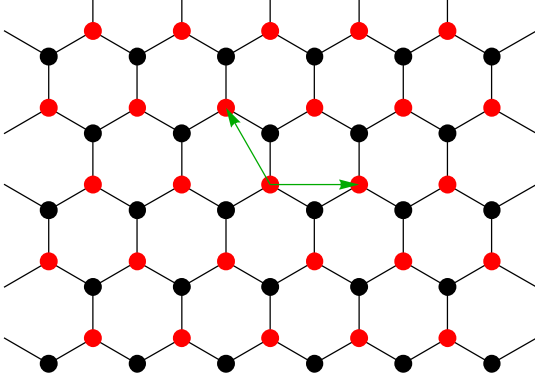
Two sites per unit cell (two sublattices) a and b (red and black respectively). The green arrow indicates the lattice vectors $\vec{v}_1$ and $\vec{v}_2$. If we shift the lattices by $m\,\vec{v}_1 + n\,\vec{v}_2$ with m and n being integers, the lattice is invariant.

$$V\!\left(\vec{r}\right) = V\!\left(\vec{r} + m\,\vec{v}_1 + n\,\vec{v}_2\right) \tag{3.307}$$

Define a to the length of the nearest-neighbor (NN) bonds, $\vec{v}_1 = \left(\sqrt{3}\,a,\,0\right)$ and $\vec{v}_2 = \left(-\sqrt{3}\,\big/2\,a,\;\;3/2\,a\right)$

If we only consider hoppings between the nearest-neighbor (NN) sites, the Hamiltonian is

$$H = -t\sum\nolimits_{<i,j>} a_i^{\dagger}\,b_j - t\sum\nolimits_{<i,j>} b_i^{\dagger}\,a_j \tag{3.308}$$

where $\langle$i,j$\rangle$ means nearest neighbors.

3.14.2. Band structures

There are three type of NN bonds: (1) along the y axis with $\theta=\pi/2$, (2) along $\theta=\pi/2+2\pi/3=7\pi/6$, (3) along $\theta=\pi/2+4\pi/3=11\pi/6$.

We need to write them out separately.

$$H = -t \sum_i a_{\vec{r}_i}^\dagger b_{r_i+\vec{e}_1} - t \sum_i a_{\vec{r}_i}^\dagger b_{r_i+\vec{e}_2} - t \sum_i a_{\vec{r}_i}^\dagger b_{r_i+\vec{e}_3} + h.c. \tag{3.309}$$

with

$$\vec{e}_1 = (0,\ a) \tag{3.310}$$

$$\vec{e}_2 = \left(-\frac{\sqrt{3}}{2}\,a,\ -\frac{a}{2} \right) \tag{3.311}$$

$$\vec{e}_3 = \left(\frac{\sqrt{3}}{2}\,a,\ -\frac{a}{2} \right) \tag{3.312}$$

Here the sum $\sum_i$ sums over all unit cells (or say all red sites).

Go to the momentum space using a 2D Fourier transformation

$$a_k = \frac{1}{\sqrt{N}} \sum_i a_i\, e^{i\vec{k}\cdot\vec{r}} \tag{3.313}$$

$$a_i = \frac{1}{\sqrt{N}} \sum_k a_k\, e^{-i\vec{k}\cdot\vec{r}} \tag{3.314}$$

$$b_k = \frac{1}{\sqrt{N}} \sum_i a_i\, e^{i\vec{k}\cdot\vec{r}} \tag{3.315}$$

$$b_i = \frac{1}{\sqrt{N}} \sum_k a_k\, e^{-i\vec{k}\cdot\vec{r}} \tag{3.316}$$

The first term in the Hamiltonian

$$-t \sum_i a_{r_i}^\dagger b_{r_i+e_1} = -\frac{t}{N} \sum_i \sum_k \sum_{k'} a_k^\dagger\, e^{i\vec{k}\cdot\vec{r}_i}\, b_{k'}\, e^{-i\vec{k}'\cdot(\vec{r}_i+\vec{e}_1)} =$$

$$-t \sum_k \sum_{k'} a_k^\dagger\, b_{k'}\, e^{-i\vec{k}'\cdot\vec{e}_1}\, \frac{1}{N} \sum_i e^{i(\vec{k}-\vec{k}')\cdot\vec{r}_i} = -t \sum_k \sum_{k'} a_k^\dagger\, b_{k'}\, e^{-i\vec{k}'\cdot\vec{e}_1}\, \delta_{k,k'} = -t \sum_k a_k^\dagger\, b_k\, e^{-i\vec{k}\cdot\vec{e}_1} \tag{3.317}$$

This is generically true. By going to the k space, one just need to change $a_{r_i}^\dagger b_{r_i+e_1}$ into $a_k^\dagger b_k$ and change $\sum_i$ (sum over all unit cells) into $\sum_k$ (sum over all momentum points), and the coefficient is $-t \exp\left[i\vec{k}\cdot(\vec{r}_a - \vec{r}_b) \right]$ with $\vec{r}_a$ and $\vec{r}_b$ being the coordinates of site a and b.

$$H = -t \sum_i a_{r_i}^\dagger b_{r_i+e_1} - t \sum_i a_{r_i}^\dagger b_{r_i+e_2} - t \sum_i a_{r_i}^\dagger b_{r_i+e_2} + h.c.$$

$$= -t \sum_k a_k^\dagger\, b_k\left(e^{-i\vec{k}\cdot\vec{e}_1} + e^{-i\vec{k}\cdot\vec{e}_2} + e^{-i\vec{k}\cdot\vec{e}_3} \right) - t \sum_k b_k^\dagger\, a_k\left(e^{i\vec{k}\cdot\vec{e}_1} + e^{i\vec{k}\cdot\vec{e}_2} + e^{i\vec{k}\cdot\vec{e}_3} \right) \tag{3.318}$$

$$= \sum_k \begin{pmatrix} a_k^\dagger & b_k^\dagger \end{pmatrix} \begin{pmatrix} 0 & \mathcal{H}_{12}(\vec{k}) \\ \mathcal{H}_{21}(\vec{k}) & 0 \end{pmatrix} \begin{pmatrix} a_k \\ b_k \end{pmatrix}$$

$$\mathcal{H}_{12}(\vec{k}) = -t\left[\exp\left(-i\vec{k}\cdot\vec{e}_1\right) + \exp\left(-i\vec{k}\cdot\vec{e}_2\right) + \exp\left(-i\vec{k}\cdot\vec{e}_3\right) \right] \tag{3.319}$$
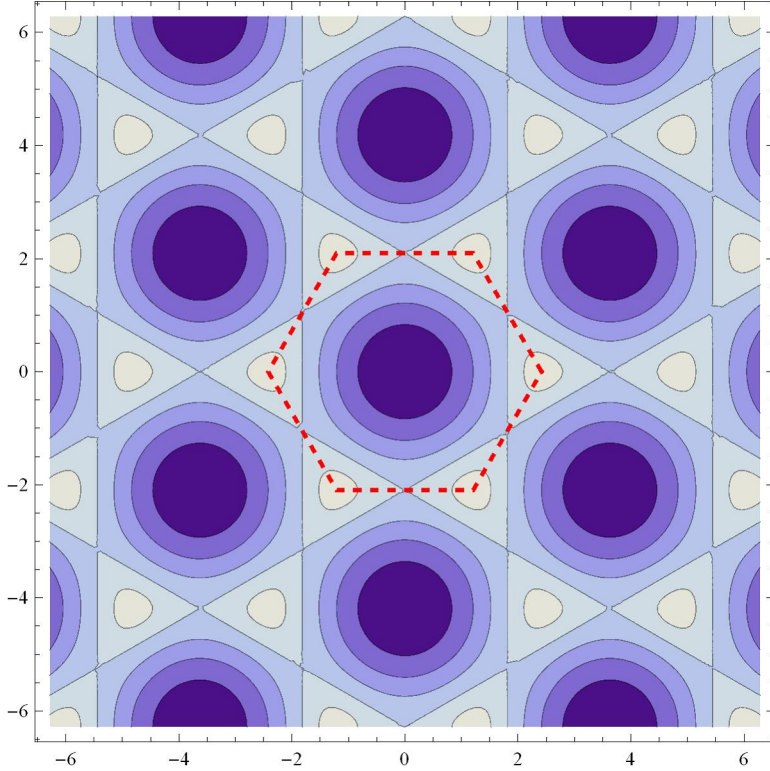
$$\mathcal{H}_{21}(\vec{k}) = \mathcal{H}_{12}(\vec{k})^* = -t\left[\exp\left(i\vec{k}\cdot\vec{e}_1\right) + \exp\left(i\vec{k}\cdot\vec{e}_2\right) + \exp\left(i\vec{k}\cdot\vec{e}_3\right) \right] \tag{3.320}$$

The kernel of the Hamiltonian:

$$\mathcal{H}\left(\vec{k}\right) = \begin{pmatrix} 0 & \mathcal{H}_{12}\left(\vec{k}\right) \\ \mathcal{H}_{21}\left(\vec{k}\right) & 0 \end{pmatrix} \tag{3.321}$$

The eigenvalues of $\mathcal{H}(k)$ gives the dispersion relation.

$$\epsilon_{\pm}\left(\vec{k}\right) = \pm \left|\mathcal{H}_{12}\left(\vec{k}\right)\right| = \pm |t| \sqrt{3 + 2\cos\left(\sqrt{3}\, k_x\, a\right) + 4\cos\left(\frac{\sqrt{3}}{2}\, k_x\, a\right)\cos\left(\frac{3}{2}\, k_y\, a\right)} \tag{3.322}$$

The contour plot of the energy dispersion for the lower band (ϵ_- as a function of k_x and k_y). The red dash lines mark the first BZ, which is a hexagon.

The dispersion $\epsilon_{\pm}$ is a periodic function of $\vec{k}$ space (the hexagon repeats itself in the figure shown above). For the lower band, its energy minimum is located at $k = 0$. The maximum of ϵ_- is reached at the corners of the BZ. The first BZ is a hexagon. It has six corners. However, due to the periodic structure in k-space, the three corner points with $\theta=0$, $\theta=2\pi/3$ and $\theta=4\pi/3$ are the same point (their momenta differ by precisely one periodicity). Similarly, the three corners with $\theta=\pi$, $\theta=\pi+2\pi/3$ and $\theta=\pi+4\pi/3$ are the same points. Therefore, there are only two different corner points and they are known as the K and K' points, where

$$K = \left(\frac{4\pi}{3\sqrt{3}\, a}, 0\right) \tag{3.323}$$

and

$$K' = \left(-\frac{4\pi}{3\sqrt{3}\, a}, 0\right) \tag{3.324}$$

For most momentum points, one of the two bands has positive energy $\epsilon_+ > 0$ and the other one has negative energy $\epsilon_- < 0$. However at the corner of the BZ, K and K', the two bands are degenerate $\epsilon_+ = \epsilon_- = 0$.

$$\epsilon_{\pm}\left(\vec{K}\right) = \pm \left|\mathcal{H}_{12}\left(\vec{K}\right)\right| = \pm \left|t\right| \sqrt{3 + 2\cos\left(\sqrt{3}\ k_x\ a\right) + 4\cos\left(\frac{\sqrt{3}}{2} k_x\ a\right)\cos\left(\frac{3}{2} k_y\ a\right)} =$$

$$\pm \left|t\right| \sqrt{3 + 2\cos\left(\sqrt{3}\ \frac{4\pi}{3\sqrt{3}\ a} a\right) + 4\cos\left(\frac{\sqrt{3}}{2}\ \frac{4\pi}{3\sqrt{3}\ a} a\right)\cos\left(\frac{3}{2} \times 0\right)} =$$

$$\pm \left|t\right| \sqrt{3 + 2\cos\left(\frac{4\pi}{3}\right) + 4\cos\left(\frac{2\pi}{3}\right)} = \pm \left|t\right| \sqrt{3 + 2\times\left(-\frac{1}{2}\right) + 4\times\left(-\frac{1}{2}\right)} = \sqrt{3-3} = 0$$

Two bands have the same energy at K and K': band crossing points.
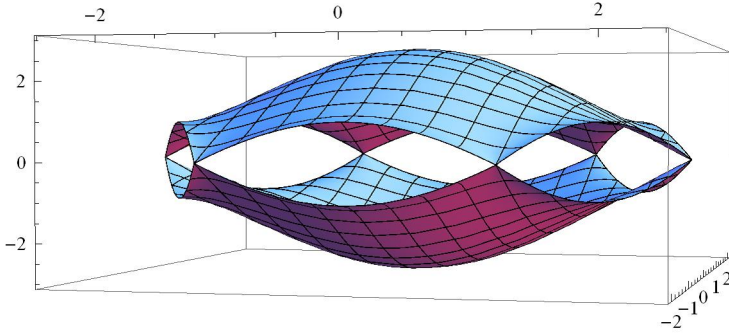
The energy dispersion for both bands ($\epsilon_{\pm}$ as a function of k_x and k_y) in the first BZ (within the red dash lines mark in the figure above).

Near K or K' points, the dispersion for the two bands are linear $\epsilon_{\pm} \propto \pm(k - K)$. To see this, we expand $\epsilon_{\pm}\left(\vec{k}\right)$ near $\vec{k} \sim \vec{K}$ and $\vec{k} \sim \vec{K}'$

$$\epsilon_{\pm}\left(\vec{k}\right) = \epsilon_{\pm}\left(\vec{K} + \vec{q}\right) = \pm \left|t\right| \sqrt{3 + 2\cos\left[\sqrt{3}\left(\frac{4\pi}{3\sqrt{3}\ a} + q_x\right)a\right] + 4\cos\left[\frac{\sqrt{3}}{2}\left(\frac{4\pi}{3\sqrt{3}\ a} + q_x\right)a\right]\cos\left(\frac{3}{2} q_y\ a\right)} =$$

$$\pm \frac{3}{2}\left|t\right|a\sqrt{q_x^2 + q_y^2} + O(q^2) = \pm \frac{3}{2}\left|t\right|a\ q + O(q^2)$$

(3.326)

$$\epsilon_{\pm}\left(\vec{k}\right) = \epsilon_{\pm}\left(\vec{K}' + \vec{q}\right) = \pm \frac{3}{2}\left|t\right|a\ q + O(q^2)$$

(3.327)
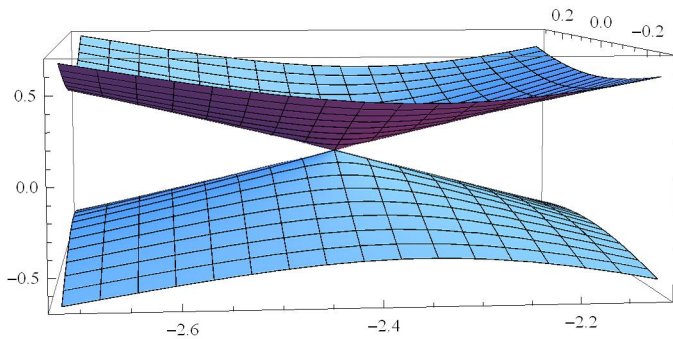
Zoom in near the K' points.

Q: Fermions with linear dispersion $\epsilon \propto k$, what is that?

A: A Massless Dirac fermion.

$$\mathcal{H}(k) = \mathcal{H}(K + q) = \frac{3}{2}\,t\,a \begin{pmatrix} 0 & q_x - i\,q_y \\ q_x + i\,q_y & 0 \end{pmatrix} + O(q^2) \approx \frac{3}{2}\,t\,a\,(q_x\,\sigma_x + q_y\,\sigma_y) = c\,\vec{q}\cdot\vec{\sigma} \tag{3.328}$$

$$\mathcal{H}(k) = \mathcal{H}(K' + q) = -\frac{3}{2}\,t\,a \begin{pmatrix} 0 & q_x + i\,q_y \\ q_x - i\,q_y & 0 \end{pmatrix} + O(q^2) \approx -\frac{3}{2}\,t\,a\,(q_x\,\sigma_x - q_y\,\sigma_y) = -c\,\vec{q}\cdot\left(\sigma_x\,\vec{\sigma}\,\sigma_x\right) \tag{3.329}$$

where c is the speed of light. Each of them is a Weyl fermion. These two Weyl fermions with opposite chirality forms a Dirac fermion.

Dirac fermions we learned in quantum mechanics:

$$i\,\partial_t\psi = \begin{pmatrix} 0 & c\,\vec{q}\cdot\vec{\sigma} \\ c\,\vec{q}\cdot\vec{\sigma} & 0 \end{pmatrix}\psi \tag{3.330}$$

If one make an unitary transformation (changing the basis, which doesn't change any physics)

$$\psi' = \begin{pmatrix} 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ \frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \\ 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ -\frac{1}{\sqrt{2}} & 0 & 0 & \frac{1}{\sqrt{2}} \end{pmatrix}\psi \tag{3.331}$$

The equation turns into

$$i\,\partial_t\psi' = \begin{pmatrix} -c\,\vec{q}\cdot\left(\sigma_x\,\vec{\sigma}\,\sigma_x\right) & 0 \\ 0 & c\,\vec{q}\cdot\vec{\sigma} \end{pmatrix}\psi' \tag{3.332}$$

The Dirac point (band crossing points) in a honeycomb lattice is stable as long as the space inversion symmetry $r \to -r$ and the time reversal symmetry $t \to -t$ are preserved. We will come back to this point latter. No matter how one perturb the systems (e.g. adding longer-range hoppings), the Dirac point is always there as long as the two symmetries mentioned above are preserved.

For graphene, the lower band is filled and the upper band is empty, which is known as "half-filling". The "half" here means that the number of electrons N_e over the number of sites N_s is $N_e/N_s = 1/2$. However, it is worthwhile to notice that the ratio between N_e and the number of unit cells N is 1, because there are two sites in each unit cell. So one of the two bands are totally filled.

By gating, one can turn the energy level slightly away from the middle point.

3.14.3. Aharonov–Bohm effect and complex hopping

Q: Can hopping strength t be complex?

A: Yes, and the phase can come from the Aharonov–Bohm effect in the presence of a magnetic field or spin-orbital couplings.

The Aharonov–Bohm effect:if one moves a particle around a closed contour. The phase difference between the final and initial states is proportional to the magnetic flux enclosed by the contour $\phi = e/\hbar \int\int B \cdot d\,S = e/\hbar \oint A.d\,l$

Discrete version on a lattice: Consider three sites a, b and c. The hopping strength between these three sites are t_{ab} t_{bc} and t_{ca} respectively. If a particle hops from a to b and then to c, the hopping strength around this loop is:

$$t_{ab}\,t_{bc}\,t_{ca} = \left|t_{ab}\right|e^{i\,\phi_{ab}} * \left|t_{bc}\right|e^{i\,\phi_{bc}} * \left|t_{ca}\right|e^{i\,\phi_{ca}} = \left|t_{ab}\,t_{bc}\,t_{ca}\right|e^{i\,(\phi_{ab}+\phi_{bc}+\phi_{ca})} \tag{3.333}$$

The phase picked up by the electron is:

$$\phi_{ab} + \phi_{bc} + \phi_{ca} = \frac{e}{\hbar}\int\int B\cdot d\,S \tag{3.334}$$

If B is nonzero inside the triangle formed by these three sites, the phase for these hoppings are nonzero.

Please notice that:

t_{ab} and t_{ba} has opposite phase, due to the Hermitian condition (One can also prove this using the Fermi's golden rule, which says that the tunneling amplitude from the state |f> to |i> is the complex conjugate of the tuning amplitude from |i> to |f>).

$$t_{ab} = t_{ba}{}^* \tag{3.335}$$

The individual phases for t_{ab}, t_{bc} and t_{ca} have no physical meaning and they are in fact gauge dependent. Only the total phase on a loop has physical meanings.

$$\phi_{ab} = e/\hbar \int_a^b A \cdot d\,l \tag{3.336}$$

Underage transformation: $A \to A + \nabla \chi$

$$\phi_{ab} \to \phi_{ab}' = e/\hbar \int_a^b (A + \nabla \chi) \cdot d\,l = e/\hbar \int_a^b A \cdot d\,l + e/\hbar \int_a^b \nabla \chi \cdot d\,l = \phi_{ab} + (\chi_b - \chi_a) \frac{e}{\hbar} \tag{3.337}$$

Obviously, ϕ_{ab} is not a physical observables, since it relies on the gauge choice. However, the total phase around a loop is different. It is a loop integral of A, which is gauge independent and the physics meaning of this Integral is the magnetic flux.

$$\phi_{ab} + \phi_{bc} + \phi_{ca} =$$
$$e/\hbar \oint A \cdot d\,l \to \phi_{ab}' + \phi_{bc}' + \phi_{ca}' = e/\hbar \oint (A + \nabla \chi) \cdot d\,l = e/\hbar \oint A \cdot d\,l + e/\hbar \oint \nabla \chi \cdot d\,l = e/\hbar \oint A \cdot d\,l = \phi_{ab} + \phi_{bc} + \phi_{ca} \tag{3.338}$$

The complex hopping strength induced by B fields breaks the time-reversal symmetry because $B \to -B$ under time reversal. (in other words, under time-reversal one need to flip the sign of all these phases).

3.14.4. Complex next-nearest-neighbor (NNN) hoppings (breaking T-symmetry using *B* fields)

Ref: F. D. M. Haldane, Model for a Quantum Hall Effect without Landau Levels: Condensed-Matter Realization of the "Parity Anomaly", Phys. Rev. Lett. 61, 2015–2018 (1988).

Now let us add some NNN hoppings and assume their hopping amplitudes are complex. For simplicity, we choose the amplitude and the complex phase to be the same for all NNN bonds. If the hopping is along the arrows marked in the figure, the phase of the hopping strength is ϕ. If one hops in the opposite direction, the phase is $-\phi$.
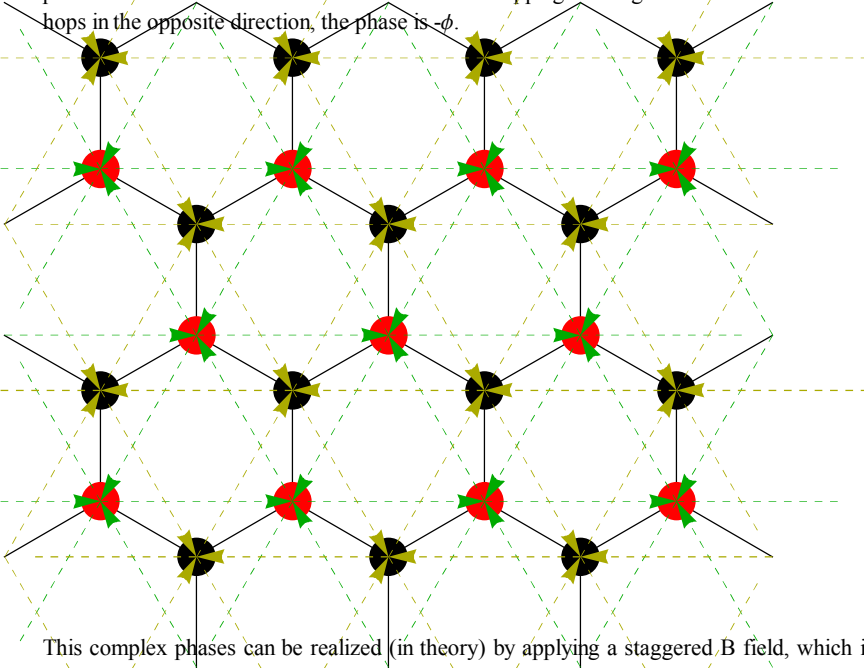
This complex phases can be realized (in theory) by applying a staggered B field, which is positive near the center of each hexagon and negative near the edges.

The NNN hoppings are from an *a*-site to an *a*-site (and from a *b*-site to a *b*-site). For a-to-a hoppings, there are three different types of NNN bonds, along $\theta=0$, $2\pi/3$ and $4\pi/3$. Same is true for b-to-b hoppings. So the Hamiltonian is:

$$H_{NNN} = -t' e^{i\phi} \sum_i a_{r_i}^\dagger a_{r_i + v_1} - t' e^{i\phi} \sum_i a_{r_i + v_1}^\dagger a_{r_i - v_3} - t' e^{i\phi} \sum_i a_{r_i - v_3}^\dagger a_{r_i} + h.c. + (a \to b \text{ and } \phi \to -\phi) \tag{3.339}$$

Here, v_1 and v_2 are marked on the first figure of this section with $\vec{v}_1 = \left(\sqrt{3}\ a,\ 0\right)$ and $\vec{v}_2 = \left(-\sqrt{3}\ \big/2\,a,\ 3/2\,a\right)$. $\vec{v}_2 = \left(-\sqrt{3}\ \big/2\,a,\ -3/2\,a\right)$

$$
\begin{aligned}
H_{\text{NNN}} &= -t'\,e^{i\phi}\sum_i a_{r_i}^\dagger a_{r_i+v_1} - t'\,e^{i\phi}\sum_i a_{r_i+v_1}^\dagger a_{r_i-v_3} - t'\,e^{i\phi}\sum_i a_{r_i-v_3}^\dagger a_{r_i} + h.c. + (a \to b \text{ and } \phi \to -\phi)\\
&= -t'\,e^{i\phi}\sum_k a_k^\dagger a_k\left(e^{-ik\cdot v_1} + e^{-ik\cdot v_2} + e^{-ik\cdot v_3}\right) + h.c. + (a \to b \text{ and } \phi \to -\phi)\\
&= -2\,t'\sum_k a_k^\dagger a_k\,[\cos(k\cdot v_1 - \phi) + \cos(k\cdot v_2 - \phi) + \cos(k\cdot v_3 - \phi)] - \\
&\quad\ 2\,t'\sum_k b_k^\dagger b_k\,[\cos(k\cdot v_1 + \phi) + \cos(k\cdot v_2 + \phi) + \cos(k\cdot v_3 + \phi)]
\end{aligned}
\tag{3.340}
$$

$$
H = H_{\text{NN}} + H_{\text{NNN}} = \sum_k \left(\, a_k^\dagger \ \ b_k^\dagger \,\right)\begin{pmatrix} \mathcal{H}_{11}(k) & \mathcal{H}_{12}(k) \\ \mathcal{H}_{21}(k) & \mathcal{H}_{22}(k) \end{pmatrix}\begin{pmatrix} a_k \\ b_k \end{pmatrix}
\tag{3.341}
$$

$$
\mathcal{H}_{12}(k) = -t\left(e^{-i\vec{k}\cdot\vec{e}_1} + e^{-i\vec{k}\cdot\vec{e}_2} + e^{-i\vec{k}\cdot\vec{e}_3}\right)
\tag{3.342}
$$

$$
\mathcal{H}_{21}(k) = \mathcal{H}_{12}(k)^* = -t\left(e^{i\vec{k}\cdot\vec{e}_1} + e^{i\vec{k}\cdot\vec{e}_2} + e^{i\vec{k}\cdot\vec{e}_3}\right)
\tag{3.343}
$$

$$
\mathcal{H}_{11}(k) = -2\,t'\,[\cos(k\cdot v_1 - \phi) + \cos(k\cdot v_2 - \phi) + \cos(k\cdot v_3 - \phi)]
\tag{3.344}
$$

$$
\mathcal{H}_{22}(k) = -2\,t'\,[\cos(k\cdot v_1 + \phi) + \cos(k\cdot v_2 + \phi) + \cos(k\cdot v_3 + \phi)]
\tag{3.345}
$$

If one computes the eigenvalues of $\mathcal{H}(k)$, one find that the two bands never cross with each other for any k (if t' is non-zero and ϕ is NOT an integer times π).
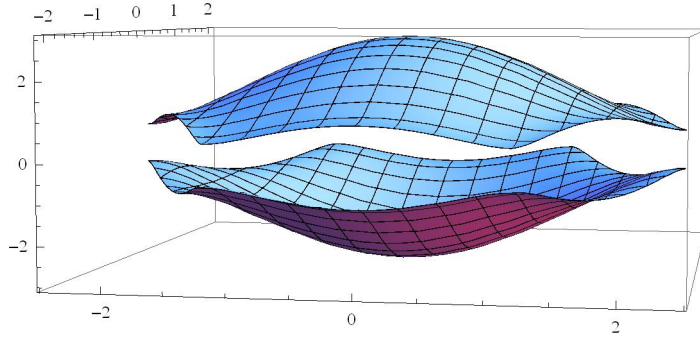
The energy dispersion for both bands ($\epsilon_\pm$ as a function of k_x and k_y) in the first BZ (within the red dash lines mark in the figure above).

Using Pauli matrices:

$$
\mathcal{H} = \mathcal{H}_0(k)\,I + \mathcal{H}_x(k)\,\sigma_x + \mathcal{H}_y(k)\,\sigma_y + \mathcal{H}_z(k)\,\sigma_z
\tag{3.346}
$$

$$
\mathcal{H}_0 = \frac{\mathcal{H}_{11}(k) + \mathcal{H}_{22}(k)}{2} = -2\,t'\cos\phi\,[\cos(k\cdot v_1) + \cos(k\cdot v_2) + \cos(k\cdot v_3)]
\tag{3.347}
$$

$$
\mathcal{H}_z = \frac{\mathcal{H}_{11}(k) - \mathcal{H}_{22}(k)}{2} = -2\,t'\sin\phi\,[\sin(k\cdot v_1) + \sin(k\cdot v_2) + \sin(k\cdot v_3)]
\tag{3.348}
$$

$$
\mathcal{H}_x = \text{Re}[\mathcal{H}_{21}(k)] = -t\left[\cos\left(\vec{k}\cdot\vec{e}_1\right) + \cos\left(\vec{k}\cdot\vec{e}_2\right) + \cos\left(\vec{k}\cdot\vec{e}_3\right)\right]
\tag{3.349}
$$

$$
\mathcal{H}_y = \text{Im}[\mathcal{H}_{21}(k)] = -t\left[\sin\left(\vec{k}\cdot\vec{e}_1\right) + \sin\left(\vec{k}\cdot\vec{e}_2\right) + \sin\left(\vec{k}\cdot\vec{e}_3\right)\right]
\tag{3.350}
$$

The energy dispersions:

$$
\epsilon_\pm\left(\vec{k}\right) = \mathcal{H}_0\left(\vec{k}\right) \pm \sqrt{\mathcal{H}_x\left(\vec{k}\right)^2 + \mathcal{H}_y\left(\vec{k}\right)^2 + \mathcal{H}_z\left(\vec{k}\right)^2}
\tag{3.351}
$$

Without the complex NNN hoppings, $\mathcal{H}_0 = \mathcal{H}_z = 0$, so the dispersion

$$
\epsilon_\pm\left(\vec{k}\right) = \pm\sqrt{\mathcal{H}_x\left(\vec{k}\right)^2 + \mathcal{H}_y\left(\vec{k}\right)^2}
\tag{3.352}
$$

At K and K', $\mathcal{H}_x = \mathcal{H}_y = 0$, so the two bands degenerate (without the complex NNN hoppings)

$$\epsilon_{\pm}\left(\vec{k}\right) = 0 \tag{3.353}$$

which gives the Dirac points.

Now, with these complex NNN hoppings, At K and K' points, $\mathcal{H}_x = \mathcal{H}_y = 0$, but $\mathcal{H}_z \neq 0$. The gap between the two bands:

$$\epsilon_{+}\left(\vec{k}\right) - \epsilon_{-}\left(\vec{k}\right) = 2\sqrt{\mathcal{H}_x\left(\vec{k}\right)^2 + \mathcal{H}_y\left(\vec{k}\right)^2 + \mathcal{H}_z\left(\vec{k}\right)^2} \tag{3.354}$$

At K or K', the gap is:

$$\epsilon_{+}\left(\vec{k}\right) - \epsilon_{-}\left(\vec{k}\right) = 2\sqrt{\mathcal{H}_x\left(\vec{k}\right)^2 + \mathcal{H}_y\left(\vec{k}\right)^2 + \mathcal{H}_z\left(\vec{k}\right)^2} = 2\left|\mathcal{H}_z\left(\vec{k} = K\right)\right| = 6\sqrt{3}\ t'\sin\phi \tag{3.355}$$

In fact, at the K point,

$$\mathcal{H}_z = -3\sqrt{3}\ t'\sin\phi \tag{3.356}$$

at the K' point

$$\mathcal{H}_z = 3\sqrt{3}\ t'\sin\phi \tag{3.357}$$

They have opposite signs (as long as ϕ is not $n\pi$).

Based on what we learned last time, this means that one cannot define the wavefunction in the whole BZ. We need to cut the systems into two regions. The region I contains the K point, the region II contains the K' points. And we need to use different eigenvectors for these two regions. Using the same method we learned in the last lecture, one finds that the Chern number here is ±1.

If one choose the boundary between region I and II to be very close to the K or K' points, one find that:

$$e^{i\,\phi(k)} = \frac{\frac{\mathcal{H}_z + \left|\vec{\mathcal{H}(k)}\right|}{\mathcal{H}_x(k) + i\,\mathcal{H}_y(k)}}{\left|\frac{\mathcal{H}_z + \left|\vec{\mathcal{H}(k)}\right|}{\mathcal{H}_x(k) + i\,\mathcal{H}_y(k)}\right|} = \frac{\left|\mathcal{H}_x(k) + i\,\mathcal{H}_y(k)\right|}{\mathcal{H}_x(k) + i\,\mathcal{H}_y(k)} = \frac{\left|q_x + i\,q_y\right|}{q_x + i\,q_y} = \frac{1}{e^{i\theta}} = e^{-i\theta} \tag{3.358}$$

3.14.5. Potential energy (breaking inversion symmetry)

Let's keep NNN hoppings to be zero for now and add some potential energy to the Hamiltonian to see whether we can open a gap to get an insulator.

$$H_{\text{Potential}} = (V + M)\sum_i a_i^\dagger a_i + (V - M)\sum_i b_i^\dagger b_i = V\,N + M\sum_i a_i^\dagger a_i - M\sum_i b_i^\dagger b_i \tag{3.359}$$

The V part (average potential between a and b sites) just adds a constant term to the energy, since the total particle number N is conserved. So we can drop the V term and only consider the difference between the potential energies at a and b sites (M).

In k-space

$$H_{\text{Potential}} = M\sum_i a_i^\dagger a_i - M\sum_i b_i^\dagger b_i = M\sum_k a_k^\dagger a_k - M\sum_k b_k^\dagger b_k = \sum_k \left(a_k^\dagger \ \ b_k^\dagger\right) M\,\sigma_z \begin{pmatrix} a_k \\ b_k \end{pmatrix} \tag{3.360}$$

It adds a σ_z component to the Hamiltonian. Same as the NNN complex hopping, this term also opens a gap at the Dirac point. If we add this term to the Hamiltonian without the complex NNN hoppings, the gap is 2M at K and K', and the system is a trivia insulator at half-filling.

Q: What will happen if we have both $H_{\text{Potential}}$ and H_{NNN}?

A: We just need to look at the signs of $\mathcal{H}_z$ at K and K' points. If they have the same sign, we can use one wavefunction to cover the whole BZ, so C=0 (trivial insulator). If they have opposite signs, the system is a topological insulator with C=±1

At the K point,

$$\mathcal{H}_z = M - 3\sqrt{3}\ t'\sin\phi \tag{3.361}$$

at the K' point

$$\mathcal{H}_z = M + 3\sqrt{3}\ t'\sin\phi \tag{3.362}$$

Therefore, as long as $\left| M \right| < \left| 3\sqrt{3}\ t'\sin\phi \right|$, the system is an topological insulator ($\mathcal{H}_z$ flips sign). If $\left| M \right| > \left| 3\sqrt{3}\ t'\sin\phi \right|$, $\mathcal{H}_z$ is always positive (or negative) and thus the system is topologically trivial.

The marginal case $| M | = \left| 3\sqrt{3}\ t'\sin\phi \right|$ is a topological transition. Here, $\mathcal{H}_z = 0$ at either the K point or the K' point. Because $\mathcal{H}_x = \mathcal{H}_y = 0$ at these two points, the gap must be zero at one of the two points. So there is a band crossing in the system (either at K o K', depending on the sign of M ant t $\sin\phi$) and the system is not an insulator.

The insulating gap is closed at the topological transition point.

Remarks:

The model of Haldane is the first example of a topological insulator beyond quantum Hall effect.

It demonstrates that topological insulator is a generic concept, which may appear in any insulating systems (NOT just quantum Hall).

It also demonstrates that as long as the topological index is nonzero, one will observe all the topological phenomena expected for a quantum Hall state, including the quantized Hall conductivity and the existence of the edge states.

The key differences between the model of Haldane and the quantum Hall effects are (1) the B field is on average zero in the model of Haldane while the QHE has a uniform B field and (2) there is a very strong lattice background in the model of Haldane while the QHE requires weak lattice potential.

Systems similar to the Haldane's model are known as topological Chern insulators or Chern insulators (average B is 0 and have a strong lattice potential). But sometimes, Chern insulator are also used to refer to the quantum Hall effect.

The model of Haldane is also the foundation to explore more complicated and exotic topological states. For example, the time-reversal invariant topological insulators was first proposed using a modified Haldane's models, which we will study latter in the semester.

$$\mathcal{H}_z = M - 3\sqrt{3}\ t'\sin\phi \tag{3.361}$$

at the K' point

$$\mathcal{H}_z = M + 3\sqrt{3}\ t'\sin\phi \tag{3.362}$$

Therefore, as long as $\left| M \right| < \left| 3\sqrt{3}\ t'\sin\phi \right|$, the system is an topological insulator ($\mathcal{H}_z$ flips sign). If $\left| M \right| > \left| 3\sqrt{3}\ t'\sin\phi \right|$, $\mathcal{H}_z$ is always positive (or negative) and thus the system is topologically trivial.

The marginal case $| M | = \left| 3\sqrt{3}\ t'\sin\phi \right|$ is a topological transition. Here, $\mathcal{H}_z = 0$ at either the K point or the K' point. Because $\mathcal{H}_x = \mathcal{H}_y = 0$ at these two points, the gap must be zero at one of the two points. So there is a band crossing in the system (either at K o K', depending on the sign of M ant t sin ϕ) and the system is not an insulator.

The insulating gap is closed at the topological transition point.

Remarks:

The model of Haldane is the first example of a topological insulator beyond quantum Hall effect.

It demonstrates that topological insulator is a generic concept, which may appear in any insulating systems (NOT just quantum Hall).

It also demonstrates that as long as the topological index is nonzero, one will observe all the topological phenomena expected for a quantum Hall state, including the quantized Hall conductivity and the existence of the edge states.

The key differences between the model of Haldane and the quantum Hall effects are (1) the B field is on average zero in the model of Haldane while the QHE has a uniform B field and (2) there is a very strong lattice background in the model of Haldane while the QHE requires weak lattice potential.

Systems similar to the Haldane's model are known as topological Chern insulators or Chern insulators (average B is 0 and have a strong lattice potential). But sometimes, Chern insulator are also used to refer to the quantum Hall effect.

The model of Haldane is also the foundation to explore more complicated and exotic topological states. For example, the time-reversal invariant topological insulators was first proposed using a modified Haldane's models, which we will study latter in the semester.

3.15. Some symmetry properties of the Berry curvature $\mathcal{F}$ and the Chern number.

3.15.1. time-reversal transformation $\mathcal{T}$ and time-reversal symmetry

In this section, we consider systems with time-reversal symmetry and shows that in the presence of the time-reversal symmetry, the Chern number must be zero. Therefore, to get a topological insulator with nonzero C, one must break the time-reversal symmetry.

Dispersion relation

Dispersion relation under time-reversal transformation:

$$\mathcal{T}\,\epsilon_n(k) = \epsilon_n(-k) \tag{3.363}$$

This is because $\mathcal{T}$ turns $\vec{k}$ into $-\vec{k}$.

There are two ways to see that $\mathcal{T}$ turns $\vec{k}$ into $-\vec{k}$.

- $\vec{k}$ is the momentum of a state. Under time-reversal, velocity changes sign, so does the momentum.

- $\vec{k} = -i\,\partial_r$. Under $\mathcal{T}$, ∂_r remains invariant. However, because $\mathcal{T}$ is an anti-unitary transformation (which changes a complex number to its complex conjugate), it flips the sign of i. Therefore, $\mathcal{T}\,\vec{k} = \mathcal{T}(-i\,\partial_r) = i\,\partial_r = -\vec{k}$.

If the system is time-reversally invariant, the time-reversal symmetry implies that $\epsilon_n(k)$ is invariant under time-reversal transformation

$$\mathcal{T}\,\epsilon_n(k) = \epsilon_n(k) \tag{3.364}$$

Compare the two equations, we have:

$$\epsilon_n(k) = \epsilon_n(-k) \tag{3.365}$$

Bottom line: For systems with time-reversal symmetry, the dispersion is an even function of the momentum $\vec{k}$.

Bloch wave function $\psi_{n,k}(r) = u_{n,k}(r) \, e^{i \, k \, r}$

Bloch wave under time-reversal transformation:

$$\mathcal{T} \, u_{n,\,k}(r) = u_{n,-k}(r)^* \tag{3.366}$$

Here, $\mathcal{T}$ changes k to $-k$ (same as the dispersion relation). In addition, because $u_{n,k}(r)$ is in general a complex function, $\mathcal{T}$ changes the function to its complex conjugate.

If the system is time-reversally invariant,

$$\mathcal{T} \, u_{n,\,k}(r) = e^{i \, \phi(k)} \, u_{n,\,k}(r) \tag{3.367}$$

By comparing the two equations above, we found that for a system with time-reversal symmetry,

$$u_{n,-k}(r)^* = e^{i \, \phi(k)} \, u_{n,\,k}(r) \tag{3.368}$$

This arbitrary phase is here because wavefunctions are defined up to an arbitrary phase

Berry connection $\mathcal{A}_n(k) = -i \, \langle u_{n,k} \mid \nabla_k \mid u_{n,k} \rangle$

The Berry connection is defined as:

$$\mathcal{A}_n(k) = -i \, \langle u_{n,k} \mid \nabla_k \mid u_{n,k} \rangle = -i \int d\vec{r} \, u_{n,k}\!\left(\vec{r}\right)^* \partial_k u_{n,k}\!\left(\vec{r}\right) \tag{3.369}$$

The under time-reversal, the new $\mathcal{A}_n(k)$ is defined as

$$\mathcal{T}\mathcal{A}_n(k) = -i \, \langle \mathcal{T} \, u_{n,k} \mid \nabla_k \mid \mathcal{T} \, u_{n,k} \rangle = -i \int d\vec{r} \, u_{n,-k}\!\left(\vec{r}\right) \partial_k u_{n,-k}\!\left(\vec{r}\right)^* =$$
$$i \int d\vec{r} \, \partial_k u_{n,-k}\!\left(\vec{r}\right) u_{n,-k}\!\left(\vec{r}\right)^* = -i \int d\vec{r} \, u_{n,-k}\!\left(\vec{r}\right)^* \partial_{-k} u_{n,-k}\!\left(\vec{r}\right) = -i \, \langle u_{n,-k} \mid \nabla_{-k} \mid u_{n,-k} \rangle = \mathcal{A}_n(-k) \tag{3.370}$$

If the system has time-reversal symmetry

$$\mathcal{T}\mathcal{A}_n(k) = \mathcal{A}_n(k) + \nabla_k \, \chi(k) \tag{3.371}$$

So we have

$$\mathcal{A}_n(-k) = \mathcal{A}_n(k) + \nabla_k \, \chi(k) \tag{3.372}$$

Bottom line: For systems with time-reversal symmetry, $\mathcal{A}_n \, (-k)$ and $\mathcal{A}_n \, (k)$ differs by a gauge transformation $\nabla_k \, \chi(k)$.

Berry curvature $\mathcal{F}_n(k) = -i\epsilon_{ij} \, \langle \partial_i u_{n,k} \mid \partial_j u_{n,k} \rangle$

The Berry curvature is defined as:

$$\mathcal{F}_n(k) = -i\epsilon_{ij} \, \langle \partial_{ki} u_{n,k} \mid \partial_{kj} u_{n,k} \rangle = -i \, \epsilon_{ij} \int d\vec{r} \, \partial_{ki} u_{n,k}\!\left(\vec{r}\right)^* \partial_{kj} u_{n,k}\!\left(\vec{r}\right) \tag{3.373}$$

The under time-reversal, the new $\mathcal{F}_n(k)$ is defined as

$$\mathcal{T}\mathcal{F}_n(k) = -i\epsilon_{ij} \, \langle \partial_{ki} \mathcal{T} u_{n,k} \mid \partial_{kj} \mathcal{T} u_{n,k} \rangle =$$
$$-i \, \epsilon_{ij} \int d\vec{r} \, \partial_{ki} \mathcal{T} u_{n,k}\!\left(\vec{r}\right)^* \partial_{kj} \mathcal{T} u_{n,k}\!\left(\vec{r}\right) = -i \, \epsilon_{ij} \int d\vec{r} \, \partial_{ki} u_{n,-k}\!\left(\vec{r}\right) \partial_{kj} u_{n,-k}\!\left(\vec{r}\right)^* = -i \, \epsilon_{ij} \int d\vec{r} \, \partial_{kj} u_{n,-k}\!\left(\vec{r}\right)^* \partial_{ki} u_{n,-k}\!\left(\vec{r}\right) \tag{3.374}$$

If we swap i and j here

$$\mathcal{T}\mathcal{F}_n(k) = -i \, \epsilon_{ji} \int d\vec{r} \, \partial_{ki} u_{n,-k}\!\left(\vec{r}\right)^* \partial_{kj} u_{n,-k}\!\left(\vec{r}\right) = i \, \epsilon_{ij} \int d\vec{r} \, \partial_{ki} u_{n,-k}\!\left(\vec{r}\right)^* \partial_{kj} u_{n,-k}\!\left(\vec{r}\right) = -\mathcal{F}_n(-k) \tag{3.375}$$

If the system has time-reversal symmetry

$$\mathcal{T}\mathcal{F}_n(k) = \mathcal{F}_n(k) \tag{3.376}$$

So we have

$$\mathcal{F}_n(-k) = -\mathcal{F}_n(k) \tag{3.377}$$

Bottom line: For systems with time-reversal symmetry, $\mathcal{F}_n(k)$ is an odd function of k

A direct result: Because the integral of an odd function over the whole Brillouin zone must be zero, (the contribution from k and -k cancels each other), the Chern number for a time-reversally invariant system must be C=0.

The Chern number $C = \frac{1}{2\pi} \iint d\,k \mathcal{F}_n(k)$

$$\mathcal{T}C = \mathcal{T} \frac{1}{2\pi} \iint d k_x \, d k_y \, \mathcal{F}_n(k) = -\frac{1}{2\pi} \iint d k_x \, d k_y \, \mathcal{F}_n(-k) = -\frac{1}{2\pi} \iint d k_x \, d k_y \, \mathcal{F}_n(k) = -C \tag{3.378}$$

If the system has time-reversal symmetry,

$$\mathcal{T}C = C, \tag{3.379}$$

So we have C=-C, which means C=0.

Bottom line: to have a nontrivial Chern number, the system must break the time-reversal symmetry (using external B field or some other method).

3.15.2. Space-inversion transformation $\mathcal{I}$ and space-inversion symmetry

Space inversion transformation $\mathcal{I}$ turns a vector $\vec{r}$ into $-\vec{r}$. It also changes $\vec{k}$ into $-\vec{k}$, which is similar to $\mathcal{T}$. However, $\mathcal{I}$ is an unitary transformation, while $\mathcal{T}$ is anti-unitary.

Dispersion relation

Dispersion relation under time-reversal transformation:

$$\mathcal{I}\epsilon_n(k) = \epsilon_n(-k) \tag{3.380}$$

If the system is invariant under space inversion

$$\mathcal{I}\,\epsilon_n(k) = \epsilon_n(k) \tag{3.381}$$

Compare the two equations, we have:

$$\epsilon_n(k) = \epsilon_n(-k) \tag{3.382}$$

Bottom line: For systems invariant under space inversion, the dispersion is an even function of the momentum $\vec{k}$.

If one want make a system where the dispersion is NOT an even function, one need to break both space inversion symmetry and time-reversal symmetry.

Bloch wave function $\psi_{n,k}(r) = u_{n,k}(r)\, e^{i\,k\,r}$

$$\mathcal{I}\,u_{n,\,k}(r) = u_{n,-k}(r) \tag{3.383}$$

If the system is invariant under space inversion

$$\mathcal{I}\,u_{n,\,k}(r) = e^{i\,\phi(k)}\,u_{n,\,k}(r) \tag{3.384}$$

By comparing the two equations above, we found that for a system with time-reversal symmetry,

$$u_{n,-k}(r) = e^{i\,\phi(k)}\,u_{n,\,k}(r) \tag{3.385}$$

Berry connection $\mathcal{A}_n(k) = -i \langle u_{n,k} \mid \nabla_k \mid u_{n,k} \rangle$

$$\mathcal{I}\mathcal{A}_n(k) = \mathcal{A}_n(-k) \tag{3.386}$$

If the system is invariant under space inversion

$$\mathcal{I}\mathcal{A}_n(k) = \mathcal{A}_n(k) + \nabla_k \chi(k) \tag{3.387}$$

So we have

$$\mathcal{A}_n(-k) = \mathcal{A}_n(k) + \nabla_k \chi(k) \tag{3.388}$$

Bottom line: For systems invariant under space inversion, $\mathcal{A}_n(-k)$ and $\mathcal{A}_n(k)$ differs by a gauge transformation $\nabla_k \chi(k)$.

Berry curvature $\mathcal{F}_n(k) = -i\epsilon_{ij}\langle \partial_i u_{n,k} \mid \partial_j u_{n,k}\rangle$

$$\mathcal{I}\mathcal{F}_n(k) = \mathcal{F}_n(-k) \tag{3.389}$$

If the system is invariant under space inversion

$$\mathcal{I}\mathcal{F}_n(k) = \mathcal{F}_n(k) \tag{3.390}$$

So we have

$$\mathcal{F}_n(-k) = \mathcal{F}_n(k) \tag{3.391}$$

Bottom line: For systems invariant under space inversion, $\mathcal{F}_n(k)$ is an even function of k

Time-reversal symmetry tell us $\mathcal{F}_n(k)$ is an odd function of k. If system have both $\mathcal{T}$ and $\mathcal{I}$ symmetry, $\mathcal{F}$ must be both an even function and an odd function. So $\mathcal{F}=0$ at any k point.

The Chern number $C = \frac{1}{2\pi} \iint d\,k \mathcal{F}_n(k)$

$$\mathcal{I}C = C \tag{3.392}$$

3.15.3. $\mathcal{I}\,\mathcal{T}$ transformation and $\mathcal{I}\mathcal{T}$ symmetry

Under $\mathcal{I}\mathcal{T}$,

$$\mathcal{I}\mathcal{T}\mathcal{F}_n(k) = \mathcal{I}[-\mathcal{F}_n(-k)] = -\mathcal{F}_n(k) \tag{3.393}$$

If system is invariant under $\mathcal{I}\mathcal{T}$,

$$\mathcal{I}\mathcal{T}\mathcal{F}_n(k) = -\mathcal{F}_n(k) \tag{3.394}$$

So, we have

$$\mathcal{F}_n(k) = -\mathcal{F}_n(k) \tag{3.395}$$

Bottom line: For systems invariant under $\mathcal{I}\mathcal{T}$, $\mathcal{F}_n(k) = 0$ at every k.

For $\mathcal{T}$ symmetry, the integral of $\mathcal{F}$ is zero (C=0), but $\mathcal{F}$ can be nonzero at each k point.

For $\mathcal{I}\mathcal{T}$ symmetry, $\mathcal{F}$ is zero at every single k point. This is a stronger statement!

But there are exceptions. For systems with $\mathcal{I}\mathcal{T}$ symmetry, $\mathcal{F}$ can actually be nonezero.

3.16. π flux and Dirac points

3.16.1. magnetic field, Aharonov–Bohm effect and magnetic π- flux

Q: Why B changes sign under $\mathcal{T}$?

A: Many ways to see it. Here, I use the A-B effect.

In the presence of B field, Aharonov and Bohm tells us that if we move a charged-particle around a closed loop, the electron will pick up a phase, and the phase is the total magnetic field enclosed by the loop (times e/$\hbar$)

$$\phi = e/\hbar \iint B \cdot dS = e/\hbar \oint A.dl \tag{3.396}$$

Under time-reversal $e^{i\phi}$ turns into $e^{-i\phi}$, because time-reversal is anti-unitary. So the A-B phase changes sign

$$\mathcal{T}\phi = -\phi \tag{3.397}$$

In other words, the integral of B changes sign

$$\mathcal{T} \iint_D B \cdot dS = - \iint_D B \cdot dS \tag{3.398}$$

Because this equation is true for any region D, we have $B = -B$.

If a system has time-reversal symmetry, everything must remain the same before and after we flip the arrow of time, including the A-B phase

$$e^{i\phi} = e^{-i\phi} \tag{3.399}$$

Therefore, we have $\phi = -\phi$. So we have $\phi=0$ for any region D, which implies that $B = 0$.

Q: Do we really need to have $B = 0$ to preserve the time-reversal symmetry.

A: NO!

This is because a phase is only well-define up to mod 2π. The time-reversal symmetry requires $e^{i\phi} = e^{-i\phi}$, which does NOT implies $\phi = -\phi$. In fact, as long as $\phi = -\phi + 2 n \pi$, the A-B phase will be the same before and after the time-reversal transformation.

This means that to preserve the time-reversal symmetry we just need $\phi = n \pi$ while n don't need to be 0.

So we need to have $\iint_D B \cdot dS = n \pi$ for any D. This means that B can contains some delta functions.

$$B = \frac{\hbar}{e} \sum_i n_i \pi \delta(r - r_i)$$
(3.400)

For this B field, the integral of B in region D is just

$$\phi = e/\hbar \iint B \cdot dS = \left(\sum_i' n_i\right) \pi$$
(3.401)

Here the sum $\sum_i'$ is over all r_i inside D. A delta function in B is a magnetic flux, which is known as a n×π-flux. These flux doesn't break the time-reversal symmetry.

In addition, it worthwhile to point out that for the A-B phase is only well defined up to mod 2π. So all the $2n\pi$ fluxes give the same A-B phase, which is 0. All the $(2n+1)\pi$ fluxes give the same A-B phase, which is π.

3.16.2. Berry flux

We can do the same thing for the Berry curvature $\mathcal{F}(k)$. If we goes around a contour in the k-space, the phase change of the Bloch wave function around this contour is

$$\phi = \iint_D \mathcal{F} \, dk^2$$
(3.402)

This is known as the Berry phase.

Under $\mathcal{IT}$, D is invariant so

$$\mathcal{IT} \, \phi = \mathcal{IT} \iint_D \mathcal{F}(k) \, dk^2 = -\iint_D \mathcal{F}(k) \, dk^2 = -\phi$$
(3.403)

If the system is $\mathcal{IT}$ invariant, we must have $\phi = -\phi$ (but up to mod 2π).

$$\phi = -\phi + 2 n \pi$$
(3.404)

so

$$\phi = n \pi$$
(3.405)

As a result, if we have $\mathcal{IT}$ symmetry, $\iint_D \mathcal{F} \, dk^2 = n\pi$ for any region D. So $\mathcal{F}$ must be either 0 or some delta functions.

$$\mathcal{F}(k) = \sum_i n_i \pi \delta(k - k_i)$$
(3.406)

These delta functions are known as Berry fluxes. Because the Berry phase is only well defined up to mod 2π, we have in general two types of Berry fluxes: 0 and π. And a π flux is a Dirac point.

3.16.3. Dirac point

Near a Dirac point, the kernel of the Hamiltonian is:

$$\mathcal{H} = v_F \begin{pmatrix} 0 & kx - i\,ky \\ kx + i\,ky & 0 \end{pmatrix}$$
(3.407)

Using polar coordinates

$$\mathcal{H} = v_F \, k \begin{pmatrix} 0 & e^{-i\phi} \\ e^{i\phi} & 0 \end{pmatrix}$$
(3.408)

The eigenvalues for this matrix is $\epsilon_\pm = \pm v_F \, k$. The eigenvectors are:

$$u_+ = \frac{1}{\sqrt{2}} \begin{pmatrix} e^{-i\phi} \\ 1 \end{pmatrix} \text{ and } u_- = \frac{1}{\sqrt{2}} \begin{pmatrix} -e^{-i\phi} \\ 1 \end{pmatrix} \tag{3.409}$$

The Berry connection in the polar coordinates has two components, radius and angle.

$$A_{+k} = -i \langle u_+ | \partial_k | u_+ \rangle = 0 \tag{3.410}$$

$$A_{+\phi} = -\frac{i}{k} \langle u_+ | \partial_\phi | u_+ \rangle = -\frac{i}{k} \frac{1}{\sqrt{2}} \begin{pmatrix} -e^{i\phi} & 1 \end{pmatrix} \partial_\phi \frac{1}{\sqrt{2}} \begin{pmatrix} -e^{-i\phi} \\ 1 \end{pmatrix} = -\frac{i}{k} \frac{1}{2} \begin{pmatrix} e^{i\phi} & 1 \end{pmatrix} \begin{pmatrix} -i\, e^{-i\phi} \\ 1 \end{pmatrix} = -\frac{1}{2k} \tag{3.411}$$

Let's choose a small circle around the Dirac point, the flux through this circle is:

$$\chi = \int\int \mathcal{F}\, dk = \oint \mathcal{A} \cdot dk = \int_0^{2\pi} \mathcal{A}_\phi\, k\, d\phi = -\int_0^{2\pi} \frac{1}{2k} k\, d\phi = -\frac{1}{2} \int_0^{2\pi} d\phi = -\pi \tag{3.412}$$

3.16.4. Why Dirac point is so stable?

This is because the Berry fluxes are very stable. If one has $\mathcal{TI}$ symmetry, the Berry curvature must be

$$\mathcal{F}(k) = \sum_i n_i\, \pi\, \delta(k - k_i) \tag{3.413}$$

Now, if we continuously tune some control parameter in our system, $\mathcal{F}(k)$ must change continuously. However, if we have $\mathcal{TI}$ symmetry, the form of $\mathcal{F}$ cannot change, the only thing that can change continuously is the location of the delta functions.

In other words, without break the $\mathcal{TI}$ symmetry, one can only move the Dirac points around in the k-space. They cannot disappear. (However, two Dirac points can annihilate each other).

$$A_{+k} = -i \langle u_+ | \partial_k | u_+ \rangle = 0 \tag{3.410}$$

$$A_{+\phi} = -\frac{i}{k} \langle u_+ | \partial_\phi | u_+ \rangle = -\frac{i}{k} \frac{1}{\sqrt{2}} \left(-e^{i\phi} \ \ 1 \right) \partial_\phi \frac{1}{\sqrt{2}} \begin{pmatrix} -e^{-i\phi} \\ 1 \end{pmatrix} = -\frac{i}{k} \frac{1}{2} \left(e^{i\phi} \ \ 1 \right) \begin{pmatrix} -i \, e^{-i\phi} \\ 1 \end{pmatrix} = -\frac{1}{2k} \tag{3.411}$$

Let's choose a small circle around the Dirac point, the flux through this circle is:

$$\chi = \int \int \mathcal{F} \, dk = \oint \mathcal{A} \cdot dk = \int_0^{2\pi} \mathcal{A}_\phi \, k \, d\phi = -\int_0^{2\pi} \frac{1}{2k} k \, d\phi = -\frac{1}{2} \int_0^{2\pi} d\phi = -\pi \tag{3.412}$$

3.16.4. Why Dirac point is so stable?

This is because the Berry fluxes are very stable. If one has $\mathcal{TI}$ symmetry, the Berry curvature must be

$$\mathcal{F}(k) = \sum_i n_i \pi \, \delta(k - k_i) \tag{3.413}$$

Now, if we continuously tune some control parameter in our system, $\mathcal{F}(k)$ must change continuously. However, if we have $\mathcal{TI}$ symmetry, the form of $\mathcal{F}$ cannot change, the only thing that can change continuously is the location of the delta functions.

In other words, without break the $\mathcal{TI}$ symmetry, one can only move the Dirac points around in the k-space. They cannot disappear. (However, two Dirac points can annihilate each other).

3.17. Edge states (numerical calculations)

3.17.1. Why energy is a function of k_x and k_y

Consider one electron moving in an infinite 2D lattice (for simplicity, we consider a square lattice here). Because the lattice is invariant under lattice translations ($x \to x + a$ and $y \to y + a$), lattice momentum is a good quantum number (a conserved quantity). Momentum conservation means that

$$[k_x, H] = [k_y, H] = 0 \tag{3.414}$$

In addition, we know $[k_x, k_y] = 0$. With three operators which commute with each other, quantum mechanics tells us that we can find common eigenstates for all these three operators, and we can use these eigenstates as the basis for the Hilbert space. This set of basis looks like:

$$\left| \psi_{\epsilon_n, k_x, k_y} \right\rangle \tag{3.415}$$

This state is the eigenstates of H, k_x and k_y with eigenvalues ϵ_n, k_x and k_y respectively. Here, one need a sub-index (n) to distinguish different states with the same k_x and k_y, which is the band index. So we have the band structure:

$$\epsilon_n(k_x, k_y) \tag{3.416}$$

3.17.2. What if we have an finite system?

To study edge states, we must have an edge. So we cannot use an infinite 2D plan. Let's consider the simplest case here, a infinitely long stripe. We assumes that the system is infinitely long along x, but it has a finite width along y. Now, the translational symmetry along x is still preserved. But there is no translational symmetry along y (due to the existence of the edge). So k_x is still a good quantum number but k_y is NOT.

$$[k_x, H] = 0 \qquad \text{but} \qquad [k_y, H] \neq 0 \tag{3.417}$$

So we cannot define common eigenstates for H, k_x and k_y, but we can still find eigenstates for H and k_x, because they commute with each other.

$$\left| \psi_{\epsilon_m, k_x} \right\rangle \tag{3.418}$$

Here, again, we need an extra index distinguish different states with the same k_x. Here I emphasize that m is NOT just the band index n. It includes information for both n and k_y.

So now, energy is a bunch of function of k_x

$$\epsilon_m(k_x) \tag{3.419}$$

where $m = 1, 2, 3 \ldots N_S$. We have N_S functions of k_x. We can draw these N_x functions on the $k_x - \epsilon$ plane, which gives us N_S curves, which is our 1D dispersion.

3.17.3. Examples

We consider the model of Haldane for a system with infinite size along x, by finite width along y. The Hamiltonian is invariant under translations along the x axis $\left(x \to x + \sqrt{3}\ a\right)$, so the x-component of the momentum is a conserved quantity $[k_x,\ H] = 0$. So we can find common eigenstates of k_x and H. The system has two edges: one on the top, one at the bottom. We can consider this lattice as a bunch of 1D (horizontal) lines coupled together, so the width are described by the number of chains one have in this infinite stripe (N). For $\epsilon_m(k_x)$, we have m=1,2,…2 N. This factor of 2 comes from the fact that we have two sites in each unit cell for a honeycomb lattice.

First, lets consider the case without NNN hopping (graphene, with two Dirac cones). Here, for the first panel, we show $\epsilon_m(k_x)$ for a system with width N=10 (so there are $N_S = 2\ N = 20$) lines. The second panel is the same calculation for a system with width N=100 ($N_S = 2\ N = 200$). The last panel shows the bulk band structure (side view).
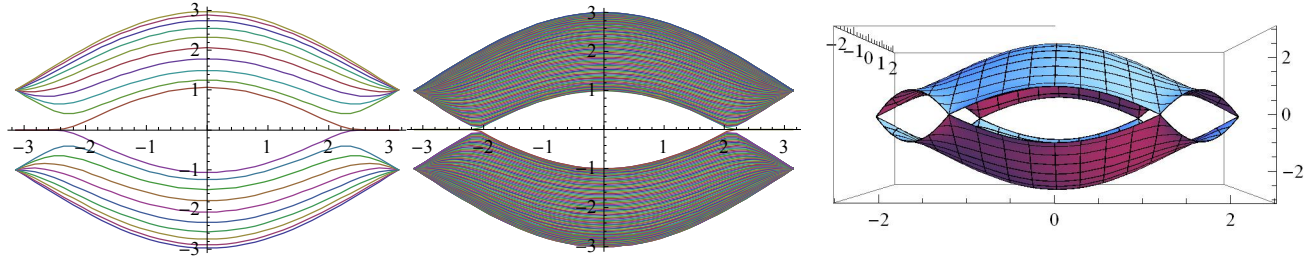
The 2D figure of $\epsilon_m(k_x)$ is in some sense the 3D band structure $\epsilon_n\left(k_x,\ k_y\right)$ projected onto a 2D plane. We can see here an upper band and a lower band and 2 Dirac cones. The system is gapless.

Now, let's consider the topological case, with complex NNN hoppings. The first panel shows the energy dispersion for an infinite stripe and the second one shows the bulk band structure for an infinite 2D system.
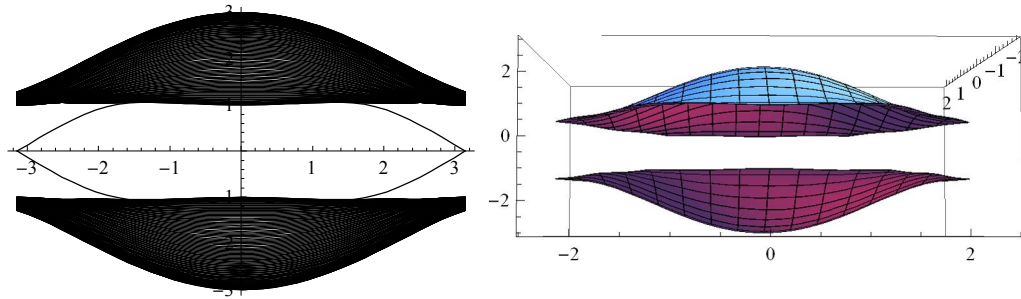
Here, I choose the parameter $\phi=\pi/2$, $t_{NN} = 1$ and $t_{NNN} = 0.2$. As one can see, for the infinite stripe, there are two bands, whose energy spectrum coincide with the bulk band structure for infinite systems. But in addition, there are two extra lines shown inside the gap, which are the metallic edge states. By looking at the wavefunctions (see the interactive figures on the course website), one find that one of the in gap state is localized near the top edge, while the other is localized near the bottom edge.

If we consider the momentum region (0,2π), instead of (-π,π), we find that one of the two edge states has positive slop and the other one has negative slop (left moving). Because the slop is the velocity of electrons ($v = \partial\epsilon/\partial k$), one of the edge states has positive velocity (right moving) while the other one is right moving (negative velocity).
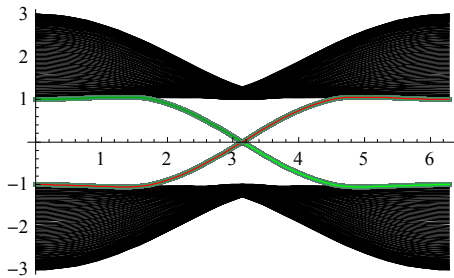
4

Gauge theory of insulators

4.1. Gauge theory (E&M) for Chern insulators

In an crystal, we have nucleons, electrons, and they couple together through gauge fields (E and B). If we don't consider the spin degree of freedom (we don't consider magnetic ordering and magnetization here), the only degrees of freedom are the motion of charge and the gauge field. In an insulator, charged particles cannot move, so the only useful degrees of freedom are the gauge fields. So let's consider the gauge field in an insulator?

4.1.1. Hamiltonian of *E* and *B* field

E and B field carries energy. In vacuum, the energy (Hamiltonian) is

$$H = \frac{1}{2} \int d\vec{r} \left(\epsilon_0 E^2 + \frac{B^2}{\mu_0} \right)$$
(4.1)

In an insulator (2D) with time-reversal symmetry, the energy of gauge fields must be a function of E and B. For weak fields (small E and B) we can do a power series expansion:

$$H = \frac{1}{2} \int d\vec{r} \left(\epsilon_x E_x^2 + \epsilon_y E_y^2 + 2 \epsilon_{xy} E_x E_y + \frac{1}{\mu_x} B_x^2 + \frac{1}{\mu_y} B_y^2 + B_z^2 \right) + \text{higher order terms}$$
(4.2)

We cannot have any other quadratic terms, like E·B, because it violates the time-reversal symmetry (B is odd under time − reversal but E is even). Higher order terms are in principle allowed and they lead to non-linear E&M response. However, these nonlinear effect are weak at small E and B, so we can ignore them.

The bottom line: the energy is still $E^2 + B^2$. Only the coefficient got renoramlized. So an insulator is NOT fundamentally different from vacuum (so we have nothing fancy here).

Q: what will happen if we break the T-reversal symmetry?

A: Additional terms are allowed by symmetry. e.g. E·B. This term means that we will induce magnetization by applying E field (and induce electric dipoles by applying B field). But this is a "small" effect, which we will not consider here. In addition to $E \cdot B$, the system can get another term which is the Chern-Simon's gauge theory

4.1.2. the Chern-Simon's gauge theory in 2+1D

For simplicity, we use the theorists' unit: $\epsilon = \mu = \text{everything} = 1$.

The Cherm-Simon's theory is easier to handle using Lagrangian, instead of the Hamiltonian. The Lagrangian is

$$L_{CS} = \frac{k}{4\pi} \int d\vec{r} \, \epsilon_{\mu\nu\lambda} A^\mu \, \partial^\nu A^\lambda$$
(4.3)

k is the coupling constant, μ, ν, λ run over t, x and y. (A^x, A^y) is the vector potential and A^t is the electric potential ϕ. $\epsilon_{txy} = \epsilon_{xyt} = \epsilon_{ytx} = 1$ and $\epsilon_{tyx} = \epsilon_{yxt} = \epsilon_{xty} = -1$ and $\epsilon = 0$ for any other subindcies.

The action is

$$S_{CS} = \int dt \, L = \frac{k}{4\pi} \int d\vec{r} \, dt \, \epsilon_{\mu\nu\lambda} \, A^\mu \, \partial^\nu \, A^\lambda \tag{4.4}$$

Gauge invariance:

Let's consider a gauge transformation:

$$A_\mu \to A_\mu' = A_\mu + \partial_\mu \chi \tag{4.5}$$

$$S_{CS} \to S_{CS}' = \frac{k}{4\pi} \int d\vec{r} \, dt \, \epsilon_{\mu\nu\lambda}(A^\mu + \partial^\mu \chi) \big[\partial^\nu \big(A^\lambda + \partial^\lambda \chi\big)\big] = \frac{k}{4\pi} \int d\vec{r} \, dt \, \epsilon_{\mu\nu\lambda}(A^\mu + \partial^\mu \chi) \, \partial^\nu \, A^\lambda =$$

$$\frac{k}{4\pi} \int d\vec{r} \, dt \, \epsilon_{\mu\nu\lambda} \, A^\mu \, \partial^\nu \, A^\lambda + \frac{k}{4\pi} \int d\vec{r} \, dt \, \epsilon_{\mu\nu\lambda} \, \partial^\mu \chi \, \partial^\nu \, A^\lambda = S_{CS} + \frac{k}{4\pi} \int d\vec{r} \, dt \, \epsilon_{\mu\nu\lambda} \, \partial^\mu \big(\chi \, \partial^\nu \, A^\lambda\big) = S_{CS} + \text{surface term} \tag{4.6}$$

If we don't worry about the surface term (assuming the system has no boundary), the action is gauge invariant. So this is a valid gauge theory. The surface term will be consider later. In fact, the surface term of the Chern-Simon's gauge theory violates the gauge symmetry. But luckily, the edge metallic state (chiral state) that we have at the edge also violates the gauge symmetry through parity anomaly. And these two effects cancels each other, so the whole system is gauge invariant.

For the whole insulator:

$$S = S_{CS} + \frac{1}{2} \int d\vec{r} \left(\epsilon_x \, E_x^2 + \epsilon_y \, E_y^2 + 2 \, \epsilon_{xy} \, E_x \, E_y + \frac{1}{\mu_x} \, B_x^2 + \frac{1}{\mu_y} \, B_y^2 + B_z^2 \right) + \text{higher order terms} \tag{4.7}$$

If the field A varies very slowly in space and time, ∂A is small. The CS term contains only one derivative, but E^2 and B^2 term contain two derivatives (E and B both take the form of ∂A). Two derivatives means it is much smaller if A changes very slowly. So here, the CS term gives the dominate contribution.

We can separate the action into two parts

$$S = S_{vacuum} + S_{material} \tag{4.8}$$

Here, $S_{material}$ is the contribution from the material (the insulator)

$$S_{material} = S - S_{vacuum} \tag{4.9}$$

Electric Current:

$$j_i(r_0, t_0) = \frac{\delta S_{material}}{\delta A_i(r_0, t_0)} = \frac{\delta}{\delta A_i(r_0, t_0)} \frac{k}{4\pi} \int d\vec{r} \, dt \, \epsilon_{\mu\nu\lambda} \, A^\mu(r, t) \, \partial^\nu \, A^\lambda(r, t) =$$

$$\frac{k}{4\pi} \int d\vec{r} \, dt \, \epsilon_{\mu\nu\lambda} \frac{\delta A^\mu(r, t)}{\delta A_i(r_0, t_0)} \, \partial^\nu \, A^\lambda(r, t) + \frac{k}{4\pi} \int d\vec{r} \, dt \, \epsilon_{\mu\nu\lambda} \, A^\mu(r, t) \Big[\partial^\nu \frac{\delta A^\lambda(r, t)}{\delta A_i(r_0, t_0)}\Big]$$

$$= \frac{k}{4\pi} \int d\vec{r} \, dt \, \epsilon_{\mu\nu\lambda} \, \delta_{\mu,i} \, \delta(r - r_0) \, \delta(t - t_0) \, \partial^\nu \, A^\lambda(r, t) + \frac{k}{4\pi} \int d\vec{r} \, dt \, \epsilon_{\mu\nu\lambda} \, A^\mu(r, t) \big[\partial^\nu \, \delta_{\lambda,i} \, \delta(r - r_0) \, \delta(t - t_0)\big]$$

$$= \frac{k}{4\pi} \, \epsilon_{i\nu\lambda} \, \partial^\nu \, A^\lambda(r_0, t_0) - \frac{k}{4\pi} \, \epsilon_{\mu\nu i} \, \partial^\nu \, A^\mu(r, t) = \frac{k}{2\pi} \, \epsilon_{i\mu\nu} \, \partial^\mu \, A^\nu(r_0, t_0) = -\frac{k}{2\pi} \, \vec{e_i} \times \vec{E_j} \tag{4.10}$$

The Hall conductivity is (we set e and $\hbar$ to 1 in the calculations above and here we add them back):

$$\sigma_{xy} = \frac{k}{2\pi} \frac{e^2}{\hbar} = k \frac{e^2}{h} \tag{4.11}$$

So σ_{xy} is quantized.

Q: Why $\frac{k}{2\pi} \, \epsilon_{i\mu\nu} \, \partial^\mu \, A^\nu(r_0, t_0) = -\frac{k}{2\pi} \, \vec{e_i} \times \vec{E_j}$?

A: let's do $i = x$ as an example:

$$j_x(r_0, t_0) = \frac{k}{2\pi} \epsilon_{x\mu\nu} \partial^\mu A^\nu(r_0, t_0) =$$

$$\frac{k}{2\pi} \epsilon_{xty} \partial^t A^y(r_0, t_0) + \frac{k}{2\pi} \epsilon_{xyt} \partial^y A^t(r_0, t_0) = \frac{k}{2\pi} \left[-\partial^t A^y(r_0, t_0) + \partial^y \phi(r_0, t_0) \right] = \frac{k}{2\pi} \left[\partial_t A^y(r_0, t_0) + \partial_y \phi(r_0, t_0) \right] = -\frac{k}{2\pi} E_y$$

(4.12)

Q: What happens if we have a metal?

A: For metals, the charge degrees of freedom must also be considered. So we have both gauge fields and charge. The situation is much more complicated and in general, we cannot get a clear answer of σ_{xy}.

$$j_x(r_0,\ t_0) = \frac{k}{2\,\pi}\,\epsilon_{x\mu\nu}\,\partial^\mu\,A^\nu(r_0,\ t_0) =$$

$$\frac{k}{2\,\pi}\,\epsilon_{xty}\,\partial^t\,A^y(r_0,\ t_0) + \frac{k}{2\,\pi}\,\epsilon_{xyt}\,\partial^y\,A^t(r_0,\ t_0) = \frac{k}{2\,\pi}\big[-\partial^t\,A^y(r_0,\ t_0) + \partial^y\,\phi(r_0,\ t_0)\big] = \frac{k}{2\,\pi}\big[\partial_t\,A^y(r_0,\ t_0) + \partial_y\,\phi(r_0,\ t_0)\big] = -\frac{k}{2\,\pi}\,E_y \tag{4.12}$$

Q: What happens if we have a metal?

A: For metals, the charge degrees of freedom must also be considered. So we have both gauge fields and charge. The situation is much more complicated and in general, we cannot get a clear answer of σ_{xy}.

4.1.3. Electric charge and magnetic field in a quantum Hall system

The Cherm-Simon's gauge theory

$$L_{CS} = \frac{k}{4\,\pi}\int d\vec{r}\,\epsilon_{\mu\nu\lambda}\,A^\mu\,\partial^\nu\,A^\lambda \tag{4.13}$$

Charge density:

$$\rho(r_0,\ t_0) = \frac{\delta S_{\text{material}}}{\delta A_0(r_0,\ t_0)} = \frac{\delta}{\delta A_0(r_0,\ t_0)}\,\frac{k}{4\,\pi}\int d\vec{r}\,dt\,\epsilon_{\mu\nu\lambda}\,A^\mu(r,\ t)\,\partial^\nu\,A^\lambda(r,\ t) =$$

$$\frac{k}{4\,\pi}\int d\vec{r}\,dt\,\epsilon_{\mu\nu\lambda}\,\frac{\delta\,A^\mu(r,\ t)}{\delta A_0(r_0,\ t_0)}\,\partial^\nu\,A^\lambda(r,\ t) + \frac{k}{4\,\pi}\int d\vec{r}\,dt\,\epsilon_{\mu\nu\lambda}\,A^\mu(r,\ t)\Big[\partial^\nu\,\frac{\delta\,A^\lambda(r,\ t)}{\delta A_0(r_0,\ t_0)}\Big]$$

$$= \frac{k}{4\,\pi}\int d\vec{r}\,dt\,\epsilon_{\mu\nu\lambda}\,\delta_{\mu,0}\,\delta(r-r_0)\,\delta(t-t_0)\,\partial^\nu\,A^\lambda(r,\ t) + \frac{k}{4\,\pi}\int d\vec{r}\,dt\,\epsilon_{\mu\nu\lambda}\,A^\mu(r,\ t)\big[\partial^\nu\,\delta_{\lambda,0}\,\delta(r-r_0)\,\delta(t-t_0)\big] \tag{4.14}$$

$$= \frac{k}{4\,\pi}\,\epsilon_{0\,\nu\lambda}\,\partial^\nu\,A^\lambda(r_0,\ t_0) - \frac{k}{4\,\pi}\,\epsilon_{\mu\nu0}\,\partial^\nu\,A^\mu(r,\ t) = \frac{k}{2\,\pi}\,\epsilon_{0\,\mu\nu}\,\partial^\mu\,A^\nu(r_0,\ t_0) = \frac{k}{2\,\pi}\,\vec{\nabla}\times\vec{A} = \frac{k}{2\,\pi}\,B$$

Bottom line: charge density is propotional to magnetic field

$$\rho(r_0,\ t_0) = \frac{k}{2\,\pi}\,B \tag{4.15}$$

Integral on both sides:

$$\iint_D dr\,\rho(r_0,\ t_0) = \frac{k}{2\,\pi}\iint_D dr\,B \tag{4.16}$$

We find that electric charge is proportional to the magnetic flux

$$q = \frac{k}{2\,\pi}\,\Phi_B \tag{4.17}$$

Flux attachment:

- Electrons in an IQH system carries magnetic flux $\Phi_B = 2\,\pi/k$ (the Chern-Simon's gauge theory attaches a magnetic flux $\Phi_B = 2\,\pi/k$ to each of the electrons).
- Magnetic flux in an IQH system carries charge q=$k\,\frac{\Phi_B}{2\,\pi}$

Physical reason: Faraday's law+Hall effect

Let's add an magnetic flux in an IQH system, by increasing the magnetic field in some region D. $B(r,\ t_i) \to B(r,\ t_f)$. If $B(t_f) > B(t_i)$, we added some magnetic flux into the system

$$\Delta\Phi_B = \iint_D dr\,B(r,\ t_f) - \iint_D dr\,B(r,\ t_i) \tag{4.18}$$

Increasing magnetic flux will create some electric field (Faraday's law)

$$\oint E\cdot dl = -\frac{d\Phi_B}{dt} \tag{4.19}$$

If we have E, it will creates Hall current perpendicular to E(Hall effect)

$$j = \sigma_{xy} E \tag{4.20}$$

Total current flowing away from the region D:

$$I = \oint j \cdot dn = \sigma_{xy} \oint E \cdot dl = -\sigma_{xy} \frac{d\Phi_B}{dt} \tag{4.21}$$

Total charge accumulated due to this current flow:

$$\Delta q = q(t_f) - q(t_i) = -\int dt\, I(t) = \sigma_{xy} \int dt\, \frac{d\Phi_B}{dt} = \sigma_{xy}\left[\Phi_B(t_f) - \Phi_B(t_i)\right] = \sigma_{xy} \Delta\Phi_B \tag{4.22}$$

If we create an flux Φ_B in a IQH system, some charge a accumulated at this flux: $q = \sigma_{xy} \Phi_B$

Chern-Simon's gauge theory: $q = k \frac{\Phi_B}{2\pi}$, adding units back $q = k \frac{e}{\hbar} \frac{\Phi_B}{2\pi} = k \frac{e}{h} \Phi_B = \sigma_{xy} \Phi_B$

4.1.4. Why k is quantized? (a not-so-rigorous proof)

(For more a rigorous proof, see Quantum Field Theory of Many-body Systems, Xiao G, Oxford 2004).

Integral over the whole system:

$$\iint_D dr\, \rho(r_0, t_0) = \frac{k}{2\pi} \iint_D dr\, B \tag{4.23}$$

Total charge is proportional to the total magnetic flux

$$N_e = k \iint_D dr\, \frac{B}{2\pi} = k\, N_m \tag{4.24}$$

N_e is always an integer. On a closed manifold, N_m is quantized to integer values (the monopole charge)

If we increase the magnetic field by a little bit such that N_m turns into $N_m + 1$, the charge of the system need to change from N_e to $N_e + k$. Because charge is always an integer, k must be an integer.

If we increase the charge by 1, $N_e \rightarrow N_e + 1$, for general k (if k is not ± 1), N_m will increase by a fractional value $1/k$ in order to satisfies the relation: $N_e = k\, N_m$. This means that if we have a IQHE with N_e electrons, by adding one electron, the system will no longer be in an IQHE effect. This is reasonable because quantum Hall effect is an insulator. For an insulator, if we add one more electron, it becomes a metal.

4.1.5. Chern-Simon's theory for an open manifold (with edges)

$$A_\mu \rightarrow A_\mu' = A_\mu + \partial_\mu \chi \tag{4.25}$$

$$S_{CS} \rightarrow S_{CS}' = S_{CS} + \frac{k}{4\pi} \int d\vec{r}\, dt\, \epsilon_{\mu\nu\lambda} \partial^\mu \left(\chi\, \partial^\nu A^\lambda\right) = S_{CS} + \text{surface term} \tag{4.26}$$

It is not gauge invariant near the boundar. This is because we have (metallic) chiral edge states near the boundary, which also violate the gauge symmetry (known as gauge anomaly), and the violation of the gauge theory there introduce an extra term to the action, which cancles the gauge dependent term here.

4.1.6. Gauge anomaly (NOT required)

Ref: Quantum theory of fields, Weinberg, VOL 2, Chapter 22, Anomalies

A stand approximation: ignore high energy states, because it cost too much energy to excite the system into such state. This approximation is NOT always valid and could be problematic.

Example (statistical physics): at 1K, what is the probability of finding a hydrogen atom in its ground state?

Energy levels of a hydrogen atom: $E = -\frac{13.6\,\text{eV}}{n^2}$

Boltzmann distribution $P \propto \exp(-E_n/T)$. Low energy state have a larger probably. For very high energy states, the probability is much smaller.

Ground state energy of a hydrogen atom: $E_1 = -13.6\,\text{eV}$

First excited states: $E_2 = -3.4\,\text{eV}$

Temperature $T = 1\ K \approx 10^{-4}$ eV.

$T << E_2 - E_1$, so let's ignore the high energy states beyond $n > 2$, only focusing on the lowest to states.

$$P_{n=1} = \frac{\exp(E_1/T)}{\exp(E_1/T) + \exp(E_2/T)} \approx \frac{10^{59\,064}}{10^{59\,064} + 10^{14\,766}} = 1 \tag{4.27}$$

This answer is WRONG! One cannot ignore high energy states. Their probably is small, but their number is huge.

$$P_{n=1} = \frac{\exp(-E_1/T)}{\sum \exp(-E_n/T)} \approx \frac{10^{59\,064}}{10^{59\,064} + 10^{14\,766} + \ldots 1 + 1 + 1 + \ldots} = \frac{10^{59\,064}}{\infty} = 0 \tag{4.28}$$

One cannot ignore the high energy states here. If one ignore the high energy states, the probably is 1. But if one include the high energy states, one get 0. In other words, the high energy states changed the probably by "-1", which is an "anomaly".

Sometimes, high energy states (ultraviolet) can change the low energy behavior (infrared) very strongly. Such a contribution from the high energy states are often called an "anomaly". Anomaly here means that one cannot see it in the low energy theory. So in the low energy theory, it seems to be something comes from nowhere. But in really, there is nothing abnormal here. Just one cannot ignore the high energy physics.

In gauge theory, there is such an effect, which was first noticed in a high-energy experiments: pion decay

pion decay: $\pi^0 \to 2\,\gamma$

Pions are bosons made by quark and anti-quark pairs.

Quarks are fermions, in 3+1D, we have left moving and right moving quarks. In the effective low energy (when we ignore the mass of fermions), the number of left and right moving quarks are separately conserved. This is one important conservation law in QED and QCD.

It is not an exact conservation law, because fermions have some mass. But it is a very good approximate conservation law.

This conservation law suppresses the pion decay rate very strongly. If we keep this (approximate) conservation law in mind, the decay rate is $1.9 \times 10^{13}\ s^{-1}$ (theory).

In experiments, the decay rate is 1.19×10^{16}, which is 1000 time larger.

The problem is the conservation law. At low energy, (if we ignore all the high energy states), there is such an conservation law (particle numbers in the left and right moving sectors are conserved separately). But the high energy state cannot be ignored in this system and destroys this conservation law, which increases the decay rate of π^0 by 3 order of magnitudes, which is known as the chiral anomaly.

A baby version of the chiral anomaly,

Let's consider a 1D condensed matter system, electrons hopping on a lattice. The dispersion is a

Low energy effective theory

$$H = \sum_k v_F k\, L_k^\dagger L_k - \sum_k v_F k\, R_k^\dagger R_k \tag{4.29}$$

We have two species of fermions, L and R. Their dispersions are $\pm v_F k$ (left and right moving).

Particle numbers for left moving particles:

$$N_L = \sum_k L_k^\dagger L_k \tag{4.30}$$

Particle numbers for right moving particles:

$$N_R = \sum_k R_k^\dagger R_k \tag{4.31}$$

N_L and N_R are both conserved quantities

$$[N_L, H] = [N_R, H] = 0 \tag{4.32}$$

This is NOT true! Because if we cannot ignore high energy states here. If we taken into account the high energy states, we find that the left and right moving branches are connected at energy far away from the chemical potential (near the band bottom). So in reality, we only have one conservation law, which is the total fermion number $N_L + N_R$. $N_L - N_R$ is NOT conserved. We only have one conserved quantity. This is the reason why one get wrong answer in π decay rate, by assuming there are two conserved quantities. The two separate conservation laws for N_L and N_R are fake. It is because one ignored high energy states, which should not be ignored in this case.

In high energy, this problem is addressed in a slightly different way, using gauge theory. Because gauge field couples to the phase of fermions and phase is related to fermion numbers, it is reasonable to use gauge theory to describe the system. In high energy, they way people describe this effect is to couple the left moving particles with a gauge field, and right moving particles with another gauge field. For left moving particles, one

effect is to couple the left moving particles with a gauge field, and right moving particles with another gauge field. For left moving particles, one find that the gauge theory is NOT gauge invariant (high energy states gives an extra term to the theory, which is not gauge invariant): $\epsilon_{\mu\nu\lambda} \frac{1}{4\pi} \chi \, \partial^\mu A^\nu$ Similarly, the right moving theory is NOT gauge invariant either, and the gauge anomaly takes the same form but with opposite sign $-\frac{1}{4\pi} \chi \nabla A^0$. To keep gauge invariance, one must have one left moving channel and one right moving channel and their gauge anomaly cancels each other. So at the end of the day, only the total number is conserved, and one cannot write down a theory which conserved the number for both left and right moving particles.

For IQHE, the edge state only have one branch (left modes only). So, the high energy theory tells us that we will have a gauge anomaly and we cannot keep the gauge theory. This gauge anomaly is precisely the extra piece we get in the CS gauge theory near the boundary. And they cancel each other, keep the whole system gauge invariant.

5

Topological insulator with time-reversal symmetry

It is impossible to have quantum Hall effect without breaking the time-reversal symmetry. $\mathcal{T}\sigma_{xy} = -\sigma_{xy}$. If we want σ_{xy} to be invariant under $\mathcal{T}$, $\mathcal{T}\sigma_{xy} = \sigma_{xy}$, σ_{xy} must be zero.

Q: Can we have other type of topological insulators, in the presence of $\mathcal{T}$-symmetry? By saying topological insulator, I mean an insulator with a protected metallic edge. By protected, I mean no matter what we do, we cannot turn the metallic edge into an insulating edge.

5.1. spin-orbit coupling

In previous discussions, we ignored the spin degrees of freedom. In reality, we have spin up and spin down electrons. If we consider free electrons moving in 2D (without lattice), the Hamiltonian is:

$$H = \sum_k \frac{k^2}{2m} c_{k\uparrow}^\dagger c_{k\uparrow} + \sum_k \frac{k^2}{2m} c_{k\downarrow}^\dagger c_{k\downarrow} \tag{5.1}$$

Kinetic energy for a particle with momentum k and spin σ, $\epsilon_\sigma(k) = k^2/2m$. The kinetic energy of a particle doesn't depend on the spin degrees of freedom. So we say that: spin degrees of freedom and orbit degrees of freedom (which gives us the kinetic energy) are decoupled.

The Hamiltonian shown above is an idea case. In reality, spin motion and orbit motion are couple together, and extra terms should be added to the Hamiltonian to describe those effect.

More general cases:

Because we have two spin species, the Hamiltonian of the kinetic energy is in general a 2×2 matrix:

$$H = \sum_k \begin{pmatrix} c_{k\uparrow}^\dagger & c_{k\downarrow}^\dagger \end{pmatrix} \mathcal{H} \begin{pmatrix} c_{k\uparrow} \\ c_{k\downarrow} \end{pmatrix} \tag{5.2}$$

Because the Hamiltonian is an Hermitian operator, $\mathcal{H}$ must be a Hermitian 2×2 matrix. For any Hermitian 2×2 matrix, we can write it in terms of identity and Pauli matrices.

$$\mathcal{H} = \mathcal{H}_0 I + \mathcal{H}_x \sigma_x + \mathcal{H}_y \sigma_y + \mathcal{H}_z \sigma_z \tag{5.3}$$

(Please notice that we used the similar construction to handle a 2 band system. Although the mathematics are similar, there is one important difference in physics. There for a two-band model, we ignored the spin index and the two-by-two matrix are NOT spin Pauli matrices. They are "pseudo-spins" (to describe the two bands). But here, the Pauli matrices are the true spin Pauli matrices, and they do tell us the spin of an electron.).

For the case we consider above, $\mathcal{H}_x = \mathcal{H}_y = \mathcal{H}_z = 0$, we only have $\mathcal{H}_0 = k^2/2m$. There is no Pauli matrix in the Hamiltonian, which means that the spin degrees of freedom plays no role in the kinetic energy (i.e. spin and orbit motion decouples).

In general, the $\mathcal{H}_x$, $\mathcal{H}_y$, $\mathcal{H}_z$ can be none-zero, so spin plays an role in the kinetic energy.

One example:

$$\mathcal{H}(k) = \begin{pmatrix} \frac{k^2}{2m} + \alpha k & 0 \\ 0 & \frac{k^2}{2m} - \alpha k \end{pmatrix} \tag{5.4}$$

More realistic model:the Rashba spin-orbital coupling:

$$\mathcal{H} = \frac{k^2}{2m} I + \alpha\,(\sigma \times k)\cdot\hat{z} = \frac{k^2}{2m} I + \alpha\,\sigma_x\,k_y - \alpha\,\sigma_y\,k_x = \begin{pmatrix} \frac{k^2}{2m} & \alpha\,k_y + i\,\alpha\,k_x \\ \alpha\,k_y - i\,\alpha\,k_x & \frac{k^2}{2m} \end{pmatrix} \tag{5.5}$$

May appear in an anisotropic crystal, the interface between two materials, or the surface of a material. For the interface, because the two materials have different electron density, electron will redistribute near the interface (moving from on material to the other). This induces an E field perpendicular to the interface. Here, the coupling strength α is determined by the strength of this E field

Another example:the Dresselhaus spin-orbital coupling:

$$\mathcal{H}_{\text{Dresselhaus}} = \beta\left[\sigma_x\,k_x\!\left(k_y{}^2 - k_z{}^2\right) + \sigma_y\,k_y\!\left(k_z{}^2 - k_x{}^2\right) + \sigma_z\,k_z\!\left(k_x{}^2 - k_y{}^2\right)\right] \tag{5.6}$$

Time-reversal symmetry

Typically, spin-orbit couplings are described by terms like $\sum_{ij} c_{ij}\,\sigma_i\,k_j{}^{2n+1}$. The spin-orbit coupling PRESERVES the time-reversal symmetry. $\mathcal{T}\sigma \to -\sigma$ and $\mathcal{T}\,k \to -k$

$$\mathcal{T}\,\sigma_i\,k_j{}^{2n+1} = (-\sigma_i)\left(-k_j{}^{2n+1}\right) = (-1)^{2n+2}\,\sigma_i\,k_j{}^{2n+1} = \sigma_i\,k_j{}^{2n+1} \tag{5.7}$$

5.2. spin-orbital coupling and magnetic field

Starting from the model we considered above

$$\mathcal{H}(k) = \begin{pmatrix} \frac{k^2}{2m} + \alpha\,k & 0 \\ 0 & \frac{k^2}{2m} - \alpha\,k \end{pmatrix} =$$

$$\begin{pmatrix} \frac{1}{2m}(k^2 + 2\,\alpha\,m\,k + m^2\,\alpha^2) - m\,\alpha^2/2 & 0 \\ 0 & \frac{1}{2m}(k^2 - 2\,\alpha\,m\,k + m^2\,\alpha^2) - m\,\alpha^2/2 \end{pmatrix} = \begin{pmatrix} \frac{(k+\alpha\,m)^2}{2m} & 0 \\ 0 & \frac{(k-\alpha\,m)^2}{2m} \end{pmatrix} - m\,\alpha^2\big/2\,I \tag{5.8}$$

Q: What does this Hamiltonian looks like?

A: electrons moving in some magnetic field. But spin up and down electrons feels opposite magnetic field.

$$H = \frac{(k - e\,A)^2}{2m} \tag{5.9}$$

We can see that for spin up electrons, they are coupled to an "magnetic field" with $A_\uparrow = -\alpha\,m/e$. For spin down electrons, it is coupled to an magnetic field with $A_\downarrow = \alpha\,m/e$. $A_\uparrow = -A_\downarrow$, so $B_\uparrow = -B_\downarrow$

Notice that this is just a toy model for demonstration purposes. Strictly speaking, one cannot define a vector potential in 1D because B=∇×A, but one cannot define cross product in 1D. So one need a 2D or 3D system to really see the analogy between B field and spin-orbit couplings. And a real system can be much more complicated.

5.3. Topological insulator from spin-orbit effect

- We can get quantum Hall insulators using B field (uniform or staggered)

- Spin-orbit effect is really similar to external B field, $\vec{B}(r)$ for spin up electron and $-\vec{B}(r)$ for spin down electron.

Q: Can we use spin-orbit effect to get topological insulators?

A: Yes. But there are two things different from IQH.

Example: Haldane's model ×2=Kane-Mele Model

- We consider to copies of Haldane's model.
- One for spin up and one for spin down
- The complex hoppings for spin up and down electrons have opposite phase.
  - Spin-up: quantum Hall state with C=1

- Spin-down: quantum Hall state with C=-1
- The whole system still have C=0 (no Hall effect), but this is an insulator with metallic edge states.

Remarks:

- In the model of Haldane, the complex hopping terms was induced by a staggered magnetic field. Here, in the Kane-Melee model, the complex hopping are spin-orbit coupling terms (spin-dependent hopping).
- Seems that the spin-orbit coupling can give us TIs without breaking the time-reversal symmetry.
- But there are two things one needs to notice (which are the key discoveries Kane and Mele made):

Model of Haldane:

$$H = H_{\text{NN}} + H_{\text{NNN}} = \sum_k \begin{pmatrix} a_k^\dagger & b_k^\dagger \end{pmatrix} \begin{pmatrix} \mathcal{H}_{11}(k) & \mathcal{H}_{12}(k) \\ \mathcal{H}_{21}(k) & \mathcal{H}_{22}(k) \end{pmatrix} \begin{pmatrix} a_k \\ b_k \end{pmatrix} \tag{5.10}$$

$$\mathcal{H}_{12}(k) = -t\left(e^{-i\vec{k}\cdot\vec{e_1}} + e^{-i\vec{k}\cdot\vec{e_2}} + e^{-i\vec{k}\cdot\vec{e_3}}\right) \tag{5.11}$$

$$\mathcal{H}_{21}(k) = \mathcal{H}_{12}(k)^* = -t\left(e^{i\vec{k}\cdot\vec{e_1}} + e^{i\vec{k}\cdot\vec{e_2}} + e^{i\vec{k}\cdot\vec{e_3}}\right) \tag{5.12}$$

$$\mathcal{H}_{11}(k) = -2\,t'\,[\cos(k\cdot v_1 - \phi) + \cos(k\cdot v_2 - \phi) + \cos(k\cdot v_3 - \phi)] \tag{5.13}$$

$$\mathcal{H}_{22}(k) = -2\,t'\,[\cos(k\cdot v_1 + \phi) + \cos(k\cdot v_2 + \phi) + \cos(k\cdot v_3 + \phi)] \tag{5.14}$$

the Kane-Mele model:

$$H = \sum_k \begin{pmatrix} a_{k\uparrow}^\dagger & b_{k\uparrow}^\dagger & a_{k\downarrow}^\dagger & b_{k\downarrow}^\dagger \end{pmatrix} \begin{pmatrix} \mathcal{H}_{11}(k) & \mathcal{H}_{12}(k) & 0 & 0 \\ \mathcal{H}_{21}(k) & \mathcal{H}_{22}(k) & 0 & 0 \\ 0 & 0 & \mathcal{H}_{11}(k) & \mathcal{H}_{21}(k) \\ 0 & 0 & \mathcal{H}_{12}(k) & \mathcal{H}_{22}(k) \end{pmatrix} \begin{pmatrix} a_{k\uparrow} \\ b_{k\uparrow} \\ a_{k\downarrow} \\ b_{k\downarrow} \end{pmatrix} \tag{5.15}$$

This is an idea model, assuming there is no hybridization between spin up and down electrons (the low-left and upper-right corner of $\mathcal{H}$ are assumed to be zero). But in any real system, this is not the case. One could get extra terms which hybridize the spin up and down electrons (extra terms in the Hamiltonian). In the presence of those extra terms, whether one still have edge states or not is very unclear. By allowing various extra terms in the model, Kane and Mele found:

1. The edge states are only protected in the presence of the time-reversal symmetry. If we break the time-reversal symmetry, the edge state will disappear and the system will turns into a trivial insulator

2. The number of edge states is only well-defined mod 2. So we only have two types of insulators: no edge states (trivial) or 1 pair of edge state (topological). This is very different from IQHE, which can have any number of edge states.

Because we only have 0 or 1 pair of edge states, these insulators are known as Z_2 topological insulators.

5.4. Why Z_2?

Let's check what will happen to the edge states, when we include extra terms into the Hamiltonian.

Case I: 1 pair of edge states.

Let's consider the Kane-Mele model (or any other model with one pair of edge states). Here our spin up electron has C=1, so it has one left moving edge states $\epsilon = v_F\,k$. Spin down electrons have C=-1, so the edge states for spin down electrons are right moving $\epsilon = -v_F\,k$. We can write them together in one Hamiltonian as a 2×2 matrix.

$$\mathcal{H}(k) = \begin{pmatrix} v_F\,k & 0 \\ 0 & -v_F\,k \end{pmatrix} = v_F\,k\,\sigma_z \tag{5.16}$$

Introducing extra terms in the bulk resulting in extra terms in the edge states Hamiltonian.

Q: Can one add any extra term to open a gap for the edge states?

This question is important. If we can open the gap for the edge states, we can move the chemical potential into the gap and the edge states turns into an insulator. If the system have an insulating edge, it is a trivial insulator (edge states are no long topologically protect). But if we can prove that no matter what we do, we can NOT open a gap, the edge is always metallic, so we have a topologically protected edge state (topological insulator).

A: No, if we preserve the time-reversal symmetry. Yes, if we break the time-reversal symmetry. In other words, if the system have time-reversal symmetry, it is a topological states with protected edge states. But if we break the time-reversal symmetry, it is a trivial insulator.

First, let's break the time-reversal symmetry, by adding a $m\sigma_z$ term to the Hamiltonian and see what will happen. This term breaks the $\mathcal{T}$-symmetry. $\mathcal{T}\sigma \rightarrow -\sigma$. So $\mathcal{T} m\sigma_z \rightarrow -m\sigma_z$

$$\mathcal{H}(k) = \begin{pmatrix} v_F k & m \\ m & -v_F k \end{pmatrix} \tag{5.17}$$

Dispersion:

$$\epsilon_{\pm}(k) = \sqrt{(v_F k)^2 + m^2} \tag{5.18}$$

We have a gap in the spectrum of the edge state. The edge state is NOT gapless. If we have chemical potential inside the gap, the edge is NOT a metal. (the system is not a topological insulator).

Second, let's preserve the time-reversal symmetry and see what will happen. If we preserve the time-reversal symmetry, the most general Hamiltonian we can write down is:

$$\mathcal{H}(k) = \sum c_{n,i} k^{2n+1} \sigma_i + \sum b_n k^{2n} I \tag{5.19}$$

At $k = 0$,

$$\mathcal{H}(k = 0) = b_0 I = \begin{pmatrix} b_0 & 0 \\ 0 & b_0 \end{pmatrix} \tag{5.20}$$

Eigenvalues at k=0: $\epsilon_{\pm} = b_0$

Gap at k=0: $\Delta(k = 0) = \epsilon_+ - \epsilon_- = 0$

No gap! No matter what chemical potential we choose, the system must be a metal at the edge. (a topological insulator).

Case II: 2 pairs of edge states.

Assume the spin up state is in a quantum Hall state with C=2, and the spin down state has C=-2. We will have two left moving electrons with spin up, and two right moving electrons with spin down.

$$H = \sum_k \begin{pmatrix} c_{k\uparrow}^\dagger & d_{k\uparrow}^\dagger & c_{k\downarrow}^\dagger & d_{k\downarrow}^\dagger \end{pmatrix} \mathcal{H} \begin{pmatrix} c_{k\uparrow} \\ d_{k\downarrow} \\ c_{k\downarrow} \\ d_{k\downarrow} \end{pmatrix} \tag{5.21}$$

$$\mathcal{H}(k) = \begin{pmatrix} v_{F1} k & 0 & 0 & 0 \\ 0 & v_{F2} k & 0 & 0 \\ 0 & 0 & -v_{F1} k & 0 \\ 0 & 0 & 0 & -v_{F2} k \end{pmatrix} \tag{5.22}$$

Now, one can add one extra term,

$$\mathcal{H}(k) = \begin{pmatrix} v_{F1} k & 0 & 0 & m \\ 0 & v_{F2} k & -m & 0 \\ 0 & -m & -v_{F1} k & 0 \\ m & 0 & 0 & -v_{F2} k \end{pmatrix} \tag{5.23}$$

This Hamiltonian is time-reversally invairnt. As we will show in the next lecture, the time-reversal operator is

$$U_T = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix} \tag{5.24}$$

Under time-reversal,

$$\mathcal{H}(k) \rightarrow \mathcal{H}'(k) = \left(U_T \mathcal{H}(k) U_T^{-1} \right)^* \tag{5.25}$$

Here * means complex conjugate. It can be easily checked that the $\mathcal{H}$(k) we wrote above is invirant under time-reversal

$$\mathcal{H}(k) \rightarrow \mathcal{H}'(k) = \left(U_T \, \mathcal{H}(k) \, U_T^{-1} \right)^* = \mathcal{H}(k) \tag{5.26}$$

Here, we consider an simple case with $v_{F1} = v_{F2}$ (for $v_{F1} \neq v_{F2}$, all qualitative results remains the same). The eigenvalues are:

Two degenerate upper band: $\epsilon = \sqrt{m^2 + k^2 \, v_{F1}{}^2}$ and two degenerate lower band $\epsilon = -\sqrt{m^2 + k^2 \, v_{F1}{}^2}$.

There is a gap $\Delta = 2m$.

Generic cases

In a time-reversal invariant 2D system:

- If we have (2n+1) pairs of edge modes, one can add extra terms to gap 2n of them. But there will always be 1 pair which is gapless (topological insulator)
- If we have 2n pairs of edge modes, one can always add some extra term to gap all of them. (trivial insulator)

Connection to high energy, one pair of edge states give us a Weyl fermion. n pairs of edge states gives us n Weyl fermions. If we enforce chiral symmetry, A Weyl fermion can only get mass (open gap) but two Weyl fermion can get mass, by combining them into a Dirac fermion. If we have 2n Weyl fermions, we can pair them up into n Dirac fermions, and give all of them some mass. But if we have 2n+1 Weyl fermions, we cannot do so. We can pair 2n of them into n Dirac fermions and give them mass, but there will be one left, which is massless (i.e. gapless).

5.5. How to compute the topological index for these Z_2 topological insulators.

- More complicated than the Chern number.
- There are many different (but equivalent) ways to do so (in the order of 10).
- Will not be discussed in the lecture because it is too technical (at least to my standard).

Here, I just describe briefly one way to get the topological index:

Step 1, Write down the wavefunction (diagonalize the 2n×2n matrix Hamiltonian for an n-band model).

Step 2, Choose the phase such that $\psi_k = \psi_{-k}{}^*$, which is always possible because we have the time-reversal symmetry.

Step 3, For a topological insulator, you will find that there is no way to define a wavefunction for every momentum point k. One need to cut the BZ into different regions and use different "gauge" in different regions.

Step 4, the gauge transformation at the boundary defines a winding number (similar to the way we find the Chern number)

Step 5, If the winding number is odd: topological! If it is even, trivial.

Remark: one may choose different gauge to repeat the calculation, and by choosing a different gauge, the winding number one find may change. But the parity of the winding number (even or odd) will never change. This is very different from the quantum Hall, where the winding number is a physical observable and the value has true physical meaning. For T-invariant topological insulators, the winding number is NOT a physical observable, and its value has no physical meaning, BUT, the parity of the winding number is physical and will not change no matter how one do the calculations.

5.6. Comparison between the time-reversal invariant topological insulators (QSHE) and QHE

The time-reversal invariant 2D topological insulators are also known as the quantum spin Hall effect (QSHE).

Symmetry:

- IQHE: Must break the time-reversal symmetry. Otherwise there is no Hall effect.
- QSHE: Must preserve the time-reversal symmetry. Otherwise the edge states are not topologically protected.

Stability:

- IQHE: More stable. Once we have IQHE, as long as we don't close the gap, the edge state will always be there.
- QSHE: Less stable. If one breaks the time-reversal symmetry, the edge states can be killed.

Impurity scatterings

- IQHE: No impurity scatterings at all.
- QSHE: Impurities cannot reflect one electron, because it breaks the time-reversal symmetry. But impurities can reflect two-electrons simultaneously, which doesn't break the time-reversal symmetry. There are impurity scatterings, but it is weaker than an ordinary 1D wire.

Experimental signature:

- IQHE: Quantized Hall conductivity (perfect quantization, error bar 1/10^9).

- QSHE: Conductivity = $2\,e^2/h$ (2 comes from the fact that we have two edge states). Notice that

  - 1. this is conductivity NOT Hall conductivity.

  - 2. this quantization is much less accurate (1% error bar for clear samples with very small size, very large deviation for larger samples or dirty samples). This is because the impurity scattering here is non-zero.

Interactions?

If we ignore the interactions and consider free fermions, IQHE and QSHE has little difference (the latter is just two copies of the former). However, if we consider interacting fermions:

- IQHE: We know that all the effect remain the same in the presence of strong interactions (in addition to the free fermion band structure theory, we also have the gauge theory, Green's function theory, and flux insertion techniques, which tell us that the Hall conductivity will remain integer-valued, even if we have very strong interactions in our system).

- QSHE: We don't have full understanding about interactions: What is the gauge theory describe (maybe BF theory)? Whether the conductivity is still $2\,e^2/h$ in the presence of strong interactions?.

3D

- IQHE: Can only happen in even dimensions 2, 4, 6 ... There is no IQHE in 3D.

- QSHE: Can be generalized to 3D

Experimental signature:

- IQHE: Quantized Hall conductivity (perfect quantization, error bar 1/10^9).

- QSHE: Conductivity = $2 e^2/h$ (2 comes from the fact that we have two edge states). Notice that

  - 1. this is conductivity NOT Hall conductivity.

  - 2. this quantization is much less accurate (1% error bar for clear samples with very small size, very large deviation for larger samples or dirty samples). This is because the impurity scattering here is non-zero.

Interactions?

If we ignore the interactions and consider free fermions, IQHE and QSHE has little difference (the latter is just two copies of the former). However, if we consider interacting fermions:

- IQHE: We know that all the effect remain the same in the presence of strong interactions (in addition to the free fermion band structure theory, we also have the gauge theory, Green's function theory, and flux insertion techniques, which tell us that the Hall conductivity will remain integer-valued, even if we have very strong interactions in our system).

- QSHE: We don't have full understanding about interactions: What is the gauge theory describe (maybe BF theory)? Whether the conductivity is still $2 e^2/h$ in the presence of strong interactions?.

3D

- IQHE: Can only happen in even dimensions 2, 4, 6 ... There is no IQHE in 3D.

- QSHE: Can be generalized to 3D

5.7. time-reversal symmetry for spin-1/2 and Kramers doublet

Q: How to write down the time-reversal operator for a spin-1/2 particle?

Basic knowledge:

- Spin 1/2: described by the Pauli matrices
- Time-reversal transformation is anti-unitary
- Time-reversal transformation change the sign of spin.

Define time-reversal operator

$$\mathcal{T} = U_T \theta \tag{5.27}$$

where U_T is an unitary matrix and θ is the operator for complex conjugate. ($\theta^{-1} = \theta$)

Because the time-reversal operator flips the sign of a spin, we have

$$\mathcal{T} \sigma_x \mathcal{T}^{-1} = -\sigma_x \text{ and } \mathcal{T} \sigma_y \mathcal{T}^{-1} = -\sigma_y \text{ and } \mathcal{T} \sigma_z \mathcal{T}^{-1} = -\sigma_z \tag{5.28}$$

$$U_T \theta \sigma_x \theta U_T^{-1} = -\sigma_x \text{ and } U_T \theta \sigma_y \theta U_T^{-1} = -\sigma_y \text{ and } U_T \theta \sigma_z \theta U_T^{-1} = -\sigma_z \tag{5.29}$$

$$\theta \sigma_x \theta = \theta \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \theta = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \theta^2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \sigma_x \tag{5.30}$$

$$\theta \sigma_z \theta = \theta \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \theta = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \theta^2 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \sigma_z \tag{5.31}$$

$$\theta \sigma_y \theta = \theta \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \theta = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} \theta^2 = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix} = -\sigma_y \tag{5.32}$$

$$U_T \sigma_x U_T^{-1} = -\sigma_x \text{ and } U_T \sigma_y U_T^{-1} = \sigma_y \text{ and } U_T \sigma_z U_T^{-1} = -\sigma_z \tag{5.33}$$

To satisfies these three relations, the only unitary matrix one can write down is

$$U_T = e^{i\phi} \sigma_y \tag{5.34}$$

where ϕ is an arbitrary phase, which depends on the choice of basis. The choice of ϕ has no physical consequence and doesn't change the value of any physical observables. Typically, people fix the value of ϕ by requiring U_T to be a real matrix ($\phi = \pi/2$ or $\phi = 3\pi/2$). If we choose $\phi = \pi/2$

$$U_T = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{5.35}$$

$\mathcal{T}^2 = ?$

Q: In classical mechanics, if we flip the arrow of time twice, everything should go back to itself ($\mathcal{T}^2 = 1$). Is this also true for a quantum system?

A: For integer spins, $\mathcal{T}^2 = 1$, but for half-integer spins $\mathcal{T}^2 = -1$

Example: spin-1/2

$$U_T = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \tag{5.36}$$

$$\mathcal{T} = U_T\, \theta = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\theta \tag{5.37}$$

$$\mathcal{T}^2 = U_T\, \theta\, U_T\, \theta = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\theta\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\theta = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\theta^2 = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} = -\mathcal{I} \tag{5.38}$$

$$\mathcal{T}\,|\psi\rangle = |\psi'\rangle \tag{5.39}$$

$$\mathcal{T}^2\,|\psi\rangle = \mathcal{T}\,|\psi'\rangle = -|\psi\rangle \tag{5.40}$$

Flip the arrow of time twice, the wavefunction for a spin-1/2 particle changes sign.

We know that there is an arbitrary phase in $U_T = e^{i\phi}\,\sigma_y$. Let's know verify that the conclusion $\mathcal{T}^2 = -1$ is independent of the choice of ϕ

$$\mathcal{T}^2 = U_T\, \theta\, U_T\, \theta = e^{i\phi}\,\sigma_y\,\theta\, e^{i\phi}\,\sigma_y\,\theta = e^{i\phi}\,\sigma_y\, e^{-i\phi}\left(-\sigma_y\right)\theta^2 = -\sigma_y\,\sigma_y = -\mathcal{I} \tag{5.41}$$

Still $\mathcal{T}^2 = -1$.

Q: What will happen if $\mathcal{T}$ is an unitary transformation?

A: for unitary transformations, $O^2 = e^{2\,i\,\phi}$, which could take any value (depending on the choice of ϕ), so it makes no sense to consider the sign of

Let's drop the anti-unitary part θ, and consider the unitary operator:

$$O = U_T = e^{i\phi}\,\sigma_y \tag{5.42}$$

$$O^2 = e^{i\phi}\,\sigma_y\, e^{i\phi}\,\sigma_y = e^{2\,i\,\phi} \tag{5.43}$$

For unitary operators, the phase of O^2 is not well defined (depending on the choice of gauge), so we cannot talk about the sign of O^2.

Bottom line: the sign of O^2 can only be defined for anti-unitary operators (like $\mathcal{T}$ or the particle-hole transformation C)

Why $\mathcal{T}^2 = -1$ for spin 1/2?

This is not an rigorous proof. But what I am going to present to you shows that spin-1/2 is special and -1 is nature for particles with spin-1/2.

Q: What will happen if we rotate the spin of a spin-1/2 particle by 2π?

Classical mechanics tells us that 2π rotation is like doing nothing. But this is NOT true for quantum mechanics.

A: For spin-1/2 particles, a 2π rotation is -1.

Define the rotation operator:

$$R_z(\theta) = \exp(-i\, S_z\, \theta/\hbar) = \exp(-i\, \sigma_z\, \theta/2) = \cos\frac{\theta}{2} - i\,\sigma_z\sin\frac{\theta}{2} \tag{5.44}$$

For $\theta = 2\pi$

$$R_z(2\,\pi) = \cos\frac{2\,\pi}{2} - i\,\sigma_z\sin\frac{2\,\pi}{2} = \cos\pi - i\,\sigma_z\sin\pi = \cos\pi = -1 \tag{5.45}$$

For a spin-1/2 particle, the wavefunction $|\psi\rangle$ changes sign under 2π rotation.

$$R_z(2\,\pi) = \cos\pi - i\,\sigma_z\sin\pi = \cos\pi = -1 \tag{5.46}$$

This -1 is the same -1 in $\mathcal{T}^2$.

The degeneracy theorem of Kramers

Q: Is $\mathcal{T}^2 = -1$ important?

A: Yes. It has a very important consequence: the degeneracy theorem of Karmers

In a time-reversely invariant system, if $\mathcal{T}^2 = -1$, all the energy levels must be doubly degenerate. In other words, if we have an eigenstate of the Hamiltonian |A⟩ with energy E, there must be another state |B⟩ which has the same energy.

- True for a single electron
- True for a system with an odd number of electrons
- True for a system with an odd total number of fermions (electrons, protons, neutrons, etc.)

For a time reversal invariant system, |ψ⟩ and $\mathcal{T}|\psi\rangle$ has the same energy. If we can proof that |ψ⟩ and $\mathcal{T}|\psi\rangle$ are two different quantum states (for $\mathcal{T}^2 = -1$), it proves automatically the theorem of Kamers.

Theorem: if $\mathcal{T}^2 = -1$, $\mathcal{T}|\psi\rangle$ give a quantum state different from |ψ⟩.

Proof: Here we first assume that $\mathcal{T}|\psi\rangle$ give the same quantum state |ψ⟩ and show it leads to contradictions

Assuming $\mathcal{T}\,|\psi\rangle = e^{i\phi}\,|\psi\rangle$

$$\mathcal{T}^2\,|\psi\rangle = \mathcal{T}\,e^{i\phi}\,|\psi\rangle = e^{-i\phi}\,\mathcal{T}\,|\psi\rangle = e^{-i\phi}\,e^{i\phi}\,|\psi\rangle = \,|\psi\rangle \tag{5.47}$$

$$\mathcal{T}^2 = +1 \tag{5.48}$$

This conclusion is in direct contradiction to the assumption $\mathcal{T}^2 = -1$. Therefore, $\mathcal{T}\,|\psi\rangle \neq e^{i\phi}\,|\psi\rangle$ for any quantum state |ψ⟩.

Comment: for $\mathcal{T}^2 = +1$ (integer spin), one can still prove that |ψ⟩ and $\mathcal{T}|\psi\rangle$ have the same energy. However, there, |ψ⟩ and $\mathcal{T}|\psi\rangle$ may be the same state. If they are the same state, we can NOT say that there are two states with the same energy. For example, the ground state of a spinless (S=0) particle in a harmonic potential trap. The ground state |G⟩ has E=ℏω/2. T|G⟩=|G⟩ and |G⟩ is unique and there is no other degenerate state has energy E=ℏω/2.

Kramers doublet and time-reversal invariant topological insulators

If we apply the theorem of Kramers to the Bloch wave in an insulator, we found that for any Bloch state $\psi_{n,k,\sigma}$, there is another state $\mathcal{T}\psi_{n,k,\sigma}$ which has the same energy.

$$\mathcal{T}\psi_{n,k,\sigma} = \psi'_{n,-k,-\sigma} \tag{5.49}$$

In general, the Kramers doublet are located at different momentum point k and $-k$. However, if $k = 0$ or $k=\pi$, $-k$ is just k itself. So we found that at $k = 0$ and $k = \pi$, every single energy level is double degenerate.

If we consider mid-gap states near $k = 0$ and $k = \pi$, they are strictly degenerate at $k = 0$ and $k = \pi$, but the degeneracy is lifted away from these two points (they are known as high symmetry points). Because energy is a smooth function of k, we need to connect these mid gap states at $k = 0$ and π. There are two different ways to connect these mid-gap states (see the two figures below).

Case I (figure on the l.h.s.): The two states at $k = 0$ are connected to the same Kramers doublet at $k = \pi$.

Case II (figure on the r.h.s.)I: The two states at $k = 0$ are connected to two different Kramers doublets at $k = \pi$ (in an zigzag way).
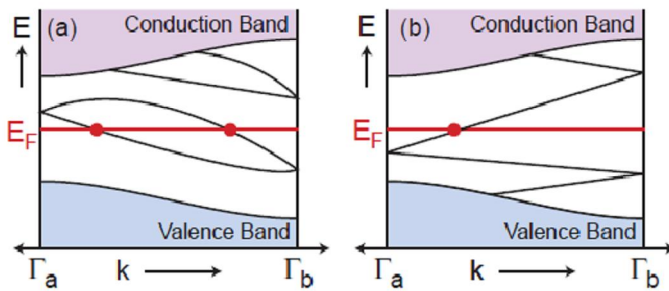

Fig. 1 the edge states

Mid-gap state for a 2D time-reversal invariant topological insulator. (figure from Topological Insulator, by Hasan and Kane, RMP 2011).

For the first case, depending on the chemical potential, we may have no edge states, two pairs of edge states, (or four pairs, or any even numbers). Because this case doesn't always have edge states, they are topologically trivial insulators.

For the second case, no matter what the value of μ is, there are always some edge states. In fact, there are always odd number of left moving edge states and the same number of right moving edge states.

Therefore, we only have two types of insulators in the presence of T-symmetry: (a) trivial insulators which have even pairs of edge states (remember that 0 is an even number) and (b) topological insulators, which have odd pairs of edge states. This is another way to understand why the Z_2 in T-invariant topological insulator.

5.8. 3D topological insulators

Time-reversal invariant topological insulator can be generalized to 3D. 3D topological insulators are described by four Z_2 topological indices. Three of them are known as weak topological indices and the last one is known as the strong topological index.
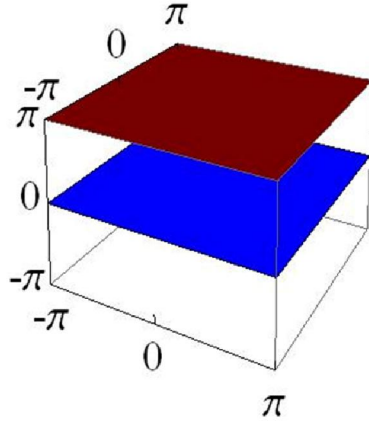
Fig. 2 the BZ of a 3D insulator

Consider the 3D BZ of a 3D insulator. Now, we consider constant k_z planes. We can consider each constant k_z plane as a 2D system. For most constant $-k_z$ planes, they are not time-revers ally invariant 2D systems, because $k_z \rightarrow -k_z$ under time-reversal. However, the $k_z = 0$ and $k_z = \pi$ planes are time-reversally invariant, because k_z is the same as $-k_z$ at 0 and π.

Now we consider $k_z = 0$ plane as a 2D time-reversal invariant insulator and $k_z = \pi$ plane as another 2D time-reversal invariant insulator. We can ask whether these two 2D planes are topological insulators or trivial insulators. There are three possibilities: both trivial, both topological, one topological and one trivial.

- Case I: both are trivial: the 3D insulator is topologically trivial
- Case II: both are nontrivial: the 3D insulator is a weak topological insulator.
- Case III: one is trivial and the other is topological: the 3D insulator is a strong topological insulator

Weak topological insulators

For a weak topological insulator, both the $k_z = 0$ and π planes have edge states. Each of them have one pair of edge states as shown in the left figure below. The dots are the Fermi surface for the edge states (where μ crosses with the dispersion relation of the edge states). However, we know that Fermi surfaces are close loops which cannot have starting or ending points. So if we have these six points of Fermi surfaces, we need to connect them to from close loops. So we get the figure on the right.
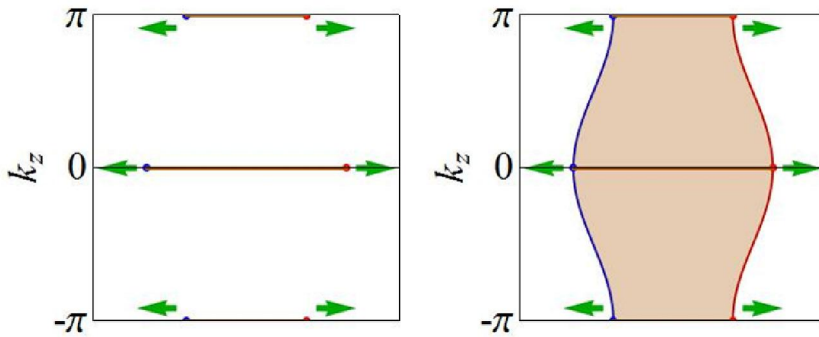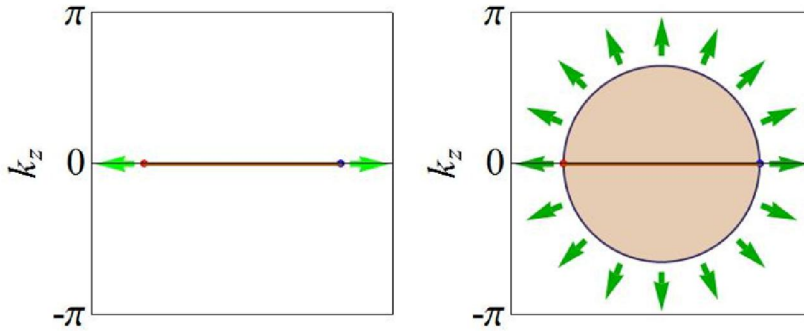

Fig. 3. the surface states of a weak topological insulator

From the figure on the right, we can see that for a 3D weak topological insulator, the 2D surface is a metal, and this surface metal has open Fermi surface as shown in the figure above.

Another way to under stand the weak topological insulator is: it is just a stack of 2D topological insulators (stacked along the z-axis).

We can repeat the same argument for constant k_x and k_y planes, which gives us two other types of weak topological insulators. They also corresponds to stacks of 2D topological insulators (but are stacked along the x or y axis.)

Strong topological insulators

For a strong topological insulator, only one of the $k_z = 0$ and π planes has edge states (say $k_z = 0$). So we have one pair of edge states as shown in the left figure below. The dots are the Fermi surface for the edge states (where μ crosses with the dispersion relation of the edge states). Again, we know that Fermi surfaces are close loops which cannot have starting or ending points. So if we have these two points of Fermi surfaces, we need to connect them to from close loops. So we get the figure on the right.

Fig. 4. the surface states of a strong topological insulator

From the figure on the right, we can see that for a 3D strong topological insulator, the 2D surface is a metal, and this surface metal has a closed Fermi surface and the spin rotate by 2π if we goes around the Fermi surface.

We can repeat the same argument for constant k_x and k_y planes. However, they don't give any new information. It we find that the system is a strong topological insulator using constant k_z planes, we will get the same answer using k_x or k_y, so there is just one types of strong topological insulator.

If we change the chemical potential, the Fermi surface changes its size (become bigger or smaller). If the Fermi surface shrink into a point, it is a Dirac point.

6

Superconductivity

6.1. Phenomena

For some metal, if one reduce the temperature, the resistivity suddenly drops to zero at some critical temperature T_C.

- Zero resistivity
- A second order phase transition at T_c (discontinuity in the specific heat)
- Meissner effect (B=0 in a superconductor).
- Magnetic field Suppresses T_C. If B is larger than some critical value B_C, the superconducting phase disappears.

6.2. Superfluid and Bose–Einstein condensate

Phenomena:

In liquid helium-4 (bosons), the viscosity (friction) vanishes below certain temperature (2.17 K).

Experiment 1: Connecting two containers with a tube and put some He4 liquid into both containers. Due to the viscosity and frictions, if we want to push the liquid He from one container to the other, we need to apply a pressure and the pressure becomes bigger if we reduce diameter of the tube. However, below the critical temperature, the fluid can flow from one side to the other without applying a pressure. This fluid with zero viscosity are known as a superfluid, which has no fiction/viscosity.

Experiment 2: Put a torsion pendulum into liquid He, at high temperature, because of the viscosity, the amplitude of the rotation decreases with time. At low temperature (below T_c), the friction and viscosity remain none zero.

Puzzle: if we measure the viscosity using experiment #1, we find it is zero in the superfulid phase. However, experiemnt #2 tells us that the viscocity is none-zero. Contradiction?

Two fluid picture:

Below T_c, the He4 liquid is a mixture of two types of fuilds: some fraction of the He atoms forms a superfuild while other He4 atoms form a normal fluid. The density for the superfulid part is called ρ_S and the density for the normal component is called ρ_N. The density of the He liquid is the sum of them $\rho = \rho_S + \rho_N$. The superfulid part has no viscosity but the normal fluid has viscosity. Above T_C, $\rho_S = 0$ and $\rho = \rho_N$. Below T_C, ρ_S increases from 0 as temperature is reduced.

In experiemnt #1, when we push the fluids, the superflluid component can flow between the two contains without any viscosity, so we always get some flow even if the pressure is very small (so we get viscocity=0). The normal fluid will not flow when the pressure is weak and don't contribute the tor flow. In other words, we not see the normal fluid in this experiment (this experiment is blind to the normal component and can only see the superfluid part).

In experiemnt #2, both the normal and superfluid components touches the pendulum. The normal fluid part induces damping. The superfluid has no friction, but the pendulum cannot see it. The pendulum only sees the normal component, which gives it damping. So we find nonezero viscosity.

Bose-Einstein condensate

For simplicy, let's ignore interactions and consider a free boson gas (in helium-4 the interactions are fairly strong and it is not a good approxiamtion to ignore interactions. However, as far as superfluidity is concerned, we can get the same phenomenon even if we ignore the interaction). The occupation number for the quantum state with momentum p is:

$$n(p) = \frac{1}{\exp\left[\left(\epsilon_p - \mu\right)\big/k_B\,T\right] - 1} \tag{6.1}$$

Here $\epsilon_p = p^2/2\,m$. Because the occupation number is nonegative $n(p) \geq 0$, $\mu < \epsilon_p$ for any p. For $\epsilon_p = p^2/2\,m$, this means that $\mu \leq 0$.

Total number of particles:

$$N = \sum_k \frac{1}{\exp\left[\left(\epsilon_k - \mu\right)/k_B\,T\right] - 1} = \frac{V}{(2\,\pi\,\hbar)^d} \int d^d\,p\,\frac{1}{\exp\left[\left(\epsilon_p - \mu\right)\big/k_B\,T\right] - 1} = \frac{V}{(2\,\pi\,\hbar)^3}\,4\,\pi \int p^2\,d\,p\,\frac{1}{\exp\left[\left(\epsilon_p - \mu\right)\big/k_B\,T\right] - 1} \tag{6.2}$$

The density of particles

$$\rho = \frac{N}{V} = \frac{1}{2\,\pi^2\,\hbar^3} \int p^2\,d\,p\,\frac{1}{\exp\left[\left(\frac{p^2}{2\,m} - \mu\right)\big/k_B\,T\right] - 1} = \frac{(2\,m\,k_B\,T)^{3/2}}{2\,\pi^2\,\hbar^3} \int x^2\,d\,x\,\frac{1}{\exp(x^2 - \mu/k_B\,T) - 1} \tag{6.3}$$

where $x = \sqrt{p/2\,m\,k_B\,T}$.

The density ρ is a function of μ and T: $\rho = \rho(\mu,\,T)$. If we plot n as a function of μ and T, we get the following figures. The l.h.s. shows a 3D plot of $\rho(\mu,\,T)$. The r.h.s. is the same function shown in a contor plot.
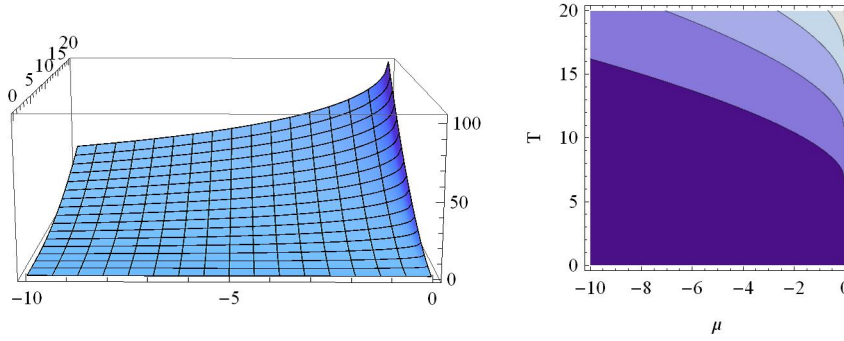

Fig. 1. Density as a function of μ and T (in arbiarty units)

From the contour plot, we can see that if we fix the density, for each temperature, we can find the corresponding chemical potential. However, the constant n curve stops at certain critical temperature T_c, at which the chemical potential reaches zero (remember that μ cannnot be positive). This means that below T_c, we cannot find a value of μ to keep the density fixed. The value of T_c is determined by the density. At T_c, $\mu=0$ so we have:

$$\rho = \frac{(2\,m\,k_B\,T_c)^{3/2}}{2\,\pi^2\,\hbar^3} \int_0^\infty x^2\,d\,x\,\frac{1}{\exp(x^2) - 1} = \frac{(2\,m\,k_B\,T_c)^{3/2}}{2\,\pi^2\,\hbar^3}\,\frac{1}{4}\,\sqrt{\pi}\;\zeta_{\frac{3}{2}} = \left(m\,k_B\,T_c/2\,\pi\,\hbar^2\right)^{3/2}\zeta_{\frac{3}{2}} \tag{6.4}$$

Here $\zeta_{3/2} = 2.61238$, with $\zeta_s = \sum_{n=1}^\infty n^{-s}$

$$T_c = \frac{2\,\pi\,\hbar^2}{m\,k_B}\,(\rho/\zeta_{3/2})^{2/3} \tag{6.5}$$

Q: What happens below T_c?

A: Bose–Einstein condensae

It turns out that the equation we have above is not entirely correct, because we used the integeral $\int d^d\,p$. For a quantum system (with finite size), we know that all quantum states are discretized, so we should use sum instead of integral. When the system size becomes larger and larger, the separation between different states becomes smaller and smaller, so that we can turn the sum into an integeral. However, in order to turn a sum into an integral, we need to assume that the function is smooth enough and has no "spikes", but this condition is violeted at and below T_c.

At T_c, we have $\mu=0$, so the occupation number is:

$$n(p) = \frac{1}{\exp\left[\frac{p^2}{2\,m\,k_B\,T}\right] - 1} \tag{6.6}$$

This function has a singularity at $p = 0$. $n(p = 0) = 1/0 = \infty$. This means that there are large number of particles in the state $p = 0$ (the ground state). However, when we compute the total number of particles, we use integral to subsituite the sum.

$$N = \sum_k \frac{1}{\exp[(\epsilon_k - \mu)/T] - 1} = \frac{V}{(2\pi\hbar)^3} 4\pi \int p^2 \, dp \, \frac{1}{\exp\left[\left(\frac{p^2}{2m} - \mu\right)/T\right] - 1} \tag{6.7}$$

In this integeral, the particles on the ground state $p = 0$ are not taken into consideration, because there is a p^2 factor in the integral. At $p = 0$, this prefactor p^2 is zreo. Therefore, using an integral is a bad approximation here, because it ignores the contribution from the ground state, which is the most important state below T_C and contains a very large number of particle (As will be shown below, above T_c it is safe to use the integral formular).

So here, we should at the ground state contribution back into the total particle number

$$N = N_0 + \frac{4\pi V}{(2\pi\hbar)^3} \int p^2 \, dp \, \frac{1}{\exp\left[\left(\frac{p^2}{2m} - \mu\right)/T\right] - 1} \tag{6.8}$$

where N_0 is the particle number in the ground state, and the integral give us the particle number on all exicated states. Similarly, for density, we have

$$\rho = \rho_0 + \frac{4\pi}{(2\pi\hbar)^3} \int p^2 \, dp \, \frac{1}{\exp\left[\left(\frac{p^2}{2m} - \mu\right)/T\right] - 1} \tag{6.9}$$

where ρ_0 is the density of the particles on the ground states and the integral gives the density for all particles on the excited states. Above T_c,

$$\rho_0 = \frac{n(p = 0)}{V} = \frac{1}{V} \frac{1}{\exp(-\mu/T) - 1} \tag{6.10}$$

In the themordynamic limit ($V \to \infty$), $\rho_0 = 0$. So we get our old integral formular back:

$$\rho = \frac{4\pi}{(2\pi\hbar)^3} \int p^2 \, dp \, \frac{1}{\exp\left[\left(\frac{p^2}{2m} - \mu\right)/T\right] - 1} \tag{6.11}$$

However, below T_c, ρ_0 cannot be ignored because:

$$\rho_0 = \frac{n(p = 0, \mu = 0)}{V} = \frac{\infty}{\infty} = ? \tag{6.12}$$

How can we determine ρ_0? We know $\rho = \rho_0 + \rho_{\text{excited}}$, so

$$\rho_0 = \rho - \rho_{\text{excited}} = \rho - \frac{4\pi}{(2\pi\hbar)^3} \int p^2 \, dp \, \frac{1}{\exp\left(\frac{p^2}{2m}/T\right) - 1} = \rho - \left(m k_B T/2\pi\hbar^2\right)^{3/2} \zeta_{\frac{3}{2}} = \left(\frac{m k_B}{2\pi\hbar^2}\right)^{3/2} \zeta_{\frac{3}{2}}\left(T_C^{3/2} - T^{3/2}\right) \tag{6.13}$$
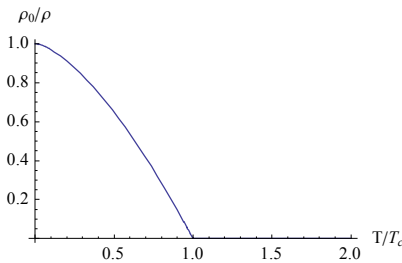
We can plot ρ_0 as a function of T.

Fig. 2. ρ_0 as a function of T.

Bose-Einstein condensate and two fluid picture

Here we seperate partices into two groups: particles in the ground state and paricles in the exictated states, and they are treated separately. This is closely related with the two-fuild picture in superfluid liquid He. ρ_0 is the superfluid part and ρ_{excited} is the normal part.

- For free bosons, BEC is a third order transition (the third order derviative of the free energy shows some discontinuity). For interacting bosons, the transition turns into a second order one.
- For free bosons, the normal component vanishes at $T = 0$. For interacting particles, there will always be some normal component left even at $T = 0$.

U(1) Phase symmetry breaking

$\rho_0 \neq 0$ is the necessary condition for superfluidilty but not surficient. We know that any phase transitio need to be associated with a symmetry breaking (unless it is a topological transition). So which symmetry is broken in the superfluid transition?

To answer this question, let's consider a theory for the "normal" component. Here, we treat the particles in the ground state as the "background" (or say as our "vaccuum"). Because we are not considering particle in the ground state, we can use the integral formulism safely (quantum field theory). However, there is a price one needs to pay for using this intergral formulas. Because we ignored the ground states, the particle conservation law is violeted. so $\langle b(r) \rangle \neq 0$.

For an operator $\hat{O}$, its expectation value is defined as the following

- At $T = 0$, $\langle \hat{O} \rangle$ is defined as $\langle G \,|\, \hat{O} \,|\, G \rangle$, where $|G\rangle$ is the ground state.
- At T>0, $\langle \hat{O} \rangle$ is defined as $\mathrm{tr}\!\left(\hat{O} \exp(\beta \hat{H})\right) = \sum_n \left\langle \psi_n \,\middle|\, \exp\!\left(\frac{\hat{H}}{k_B T}\right) \hat{O} \,\middle|\, \psi_n \right\rangle$ where $\{\psi_n\}$ is a complete basis of the Hilbert space, $\hat{H}$ is the

  Hamiltonian operator. If $|\psi_n\rangle$ are the eigenstates of H with eigenenergy E_n, $\langle \hat{O} \rangle$ becomes $\sum_n \exp(-E_n / k_B T) \langle \psi_n \,|\, O \,|\, \psi_n \rangle$, which recovers the Canonical ensemble we learned in classical statistical mechanics.

For operator $b(r)$, its expectation in the normal phase must be zero, due to the particle conservation law. However, in the superfluid phase, because we ignored the ground states, which has a large number of particles, $\langle b \rangle \neq 0$ and $\langle b \rangle \propto \sqrt{\rho_0}$.

How do we know that $\langle b \rangle \propto \sqrt{\rho_0}$? The simplest way to show it is via dimension analysis. $\int d^d r \, \langle b^\dagger(r) \, b(r) \rangle = N$ is the total number of particle. N is a dimensionliess quantity, so $\langle b^\dagger(r) \, b(r) \rangle$ must have the dimension of r^{-d} (density). So $\langle b \rangle$ must have the dimension of the square root density. So, $\langle b \rangle \propto \sqrt{\rho_0}$. Therefore, we know that $\langle b \rangle = 0$ above T_c and $\langle b \rangle \neq 0$ below T_c. So we can use $\langle b \rangle$ as our order parameter. Define $\psi(r) = \langle b(r) \rangle$ to be our order parameter, we can write down an phenomenological theory (the Ginsburg-Landau free energy):

$$F(\psi, \psi^*) = \int \gamma(T) \, \nabla \psi^* \, \nabla \psi + \alpha(T) \, \psi^* \psi + \beta(T) \, (\psi^* \psi)^2 + \ldots \tag{6.14}$$

For stablity reasons, we assume $\gamma>0$ and $\beta>0$ and $\alpha = \alpha_0(T - T_c)$. similar to what we learned from our thermaldynamics class, we need to minimze the free energy. Because $\gamma>0$, we'd better have $\nabla \psi = 0$ (uniform ψ).

$$F(\psi, \psi^*) = \int \alpha_0(T - T_c) \, (\,|\psi|\,)^2 + \beta(T) \, (\,|\psi|\,)^4 \tag{6.15}$$

For $T > T_c$, the minimum of F is reached at $\psi=0$. For $T < T_C$, the minimum of F has $|\psi| = \sqrt{\alpha_0(T_c - T)/2\,\beta}$. However, the phase of ψ can take any value. The phase is choosen "randomly". This can be seen by ploting F as a function of the real and imaginary parts of ψ, we get the famous Mexician hat free energy:
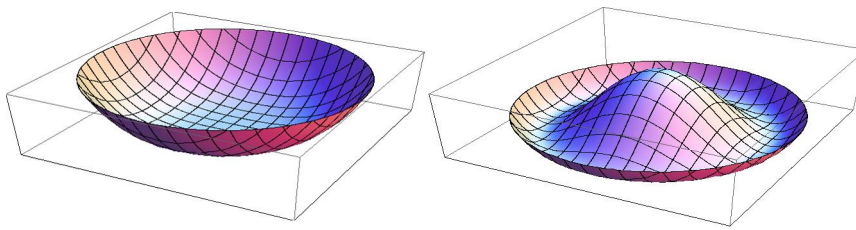

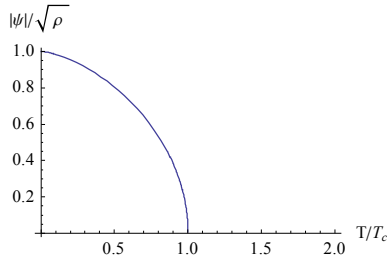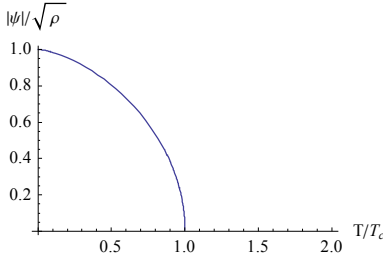Fig. 3. *F* as a function of Re(ψ) and Im(ψ)

Fig. 4. $|\psi|$ as a function of T.

For $T > T_c$, F has only one minimum with ψ=0. So we only have one ground state. However at $T < T_c$, the ψ=0 state becomes unstable and there are infinite number of minimum of F on the ring of $|\psi| = \sqrt{\alpha_0(T_c - T)/2\,\beta}$. This means that the system have infinite number of degenreate ground state. Which ground state the system falls into is determined determined by chance and they all have the same opportunity.

Here the Hamiltonian has the U(1) phase symmetry. $b \to b\, e^{i\,\phi}$. However, the ground state doesn't have this U(1) symmetry. ($\langle b \rangle$ is not invariant under $b \to b\, e^{i\,\phi}$, because $\langle b \rangle \to \langle b \rangle\, e^{i\,\phi}$). So the low temperature superfluid phase breaks the U(1) phase symmetry.

Fig. 4. $|\psi|$ as a function of T.

For $T > T_c$, F has only one minimum with $\psi=0$. So we only have one ground state. However at $T < T_c$, the $\psi=0$ state becomes unstable and there are infinite number of minimum of F on the ring of $|\psi| = \sqrt{\alpha_0(T_c - T)/2\,\beta}$. This means that the system have infinite number of degenreate ground state. Which ground state the system falls into is determined determined by chance and they all have the same opportunity.

Here the Hamiltonian has the U(1) phase symmetry. $b \to b\,e^{i\,\phi}$. However, the ground state doesn't have this U(1) symmetry. ($\langle b \rangle$ is not invariant under $b \to b\,e^{i\,\phi}$, because $\langle b \rangle \to \langle b \rangle\,e^{i\,\phi}$). So the low temperature superfluid phase breaks the U(1) phase symmetry.

6.3. Cooper Pair and the BCS theory

Superconductivity and superfluid are very similar (no resistivity vs no viscosity), which implies that the fundamental physics may be similar. But electrons are fermions, and we cannot have more than 1 electron on quantum state, so there is no way to have BEC for fermions (which requires many particles to stay on the ground state). How can fermions form a BEC state? Well, one fermion cannot condensate, but a pair of fermions is a boson and these pairs can condensate. These fermion pairs are known as Cooper pairs, and this theory of superconductivity is known as the BCS theory (for Bardeen, Cooper and Schrieffer).

Originally this idea sounds like crazy, because electrons the same charge so they have repulsive interactions. How can a bunch of particles

6.3.1. Electron phonon interactions

A rigorous treatment on electron-phonon interactions requires quantum field theory. Due to time-limitation here are just describe the physical picture without rigorous proof. Key conclusions:

- Electron-phonon interactions are crucial for conventional superconductors.
- Electron-phonon interactions induce an effective attractive interactions between electrons.

The first conclusion are based on the isotopic effect in conventional superconductors. If superconductor are purely due to electrons and lattices places no roles in it, the transition temperature Tc should be independent on the mass of the nucleons. However, experiments shows that the transition temperature T_c depends strongly on the mass of the nucleons, if we use different isotopes of a given element type of atom (e.g. Hg 203, Hg 202, Hg 200, Hg 199, Hg 198). This shows a hint that superconductor are related with the lattice motion (phonons).

The second conclusion are based on this physics picture

- An electron carries negative charge, which will attract nearby nucleons. So locally, the spacing between the nucleons will be reduced.
- Electron mass is much smaller than nucleon mass. So electrons moves much faster.
- When electron flies away, the lattice distortion will not recover immediately (because lattice moves slower). So locally, there is a higher density of positive charges here, which will attract other electrons.
- Combining all the things mentioned above, we found that an electron will attract other electrons, via lattice distortions. And this is phonon mediated attractions between electrons.

Theoretically, the attraction are induced when two electrons exchange a virtue phonon, which can shown rigorously in quantum field theory.

6.3.2. Model Hamiltonian

The key conclusion in the BCS theory is that: in a Fermi liquid, as long as there are some attractive interactions (no matter how weak the interactions are), the Fermi liquid state will become unstable below some temperature T_c.

Now, let's consider electrons with attractive interactions. The Hamiltonian is

The energy contains to parts, kinetic energy and the potential energy.

$$E = \sum_{k,\sigma} \epsilon(k)\, n_\sigma(k) + \frac{1}{2} \sum_{r,r'} \sum_{\sigma,\sigma'} V(r - r')\, n_\sigma(r)\, n_{\sigma'}(r') \tag{6.16}$$

If we translate this into the form of second quantization:

$$H = \sum_{k,\sigma} \epsilon(k)\, c_{k\sigma}^{\dagger}\, c_{k\sigma} + \frac{1}{2} \sum_{\sigma,\sigma'} \sum_{r,r'} V(r - r')\, c_{r\sigma}^{\dagger}\, c_{r\sigma}\, c_{r'\sigma'}^{\dagger}\, c_{r'\sigma'} \tag{6.17}$$

Transfer the interaction part into the k space.

$$H = \sum_{k,\sigma} \epsilon(k)\, c_{k,\sigma}^{\dagger}\, c_{k,\sigma} + \frac{1}{2} \sum_{q,k,k'} \sum_{\sigma,\sigma'} V_q\, c_{k+q/2,\sigma}^{\dagger}\, c_{k-q/2,\sigma}\, c_{k'-q/2,\sigma'}^{\dagger}\, c_{k'+q/2,\sigma'} \tag{6.18}$$

Let's simplify further the interactions:

$$H = \sum_{k,\sigma} \epsilon(k)\, c_{k,\sigma}^{\dagger}\, c_{k,\sigma} + \sum_{k,k'} V_{k,k'}\, c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger}\, c_{k'\downarrow}\, c_{-k'\uparrow} \tag{6.19}$$

This interaction is part of the electron interactions. Here we ignored all other type of interactions. This turns out to be a good approximation, because this interaction is the key which induces the pairing while all other interactions have very little effects.

For simplicity, we assume that $V_{k,k'} = V$.

$$H = \sum_{k,\sigma} \epsilon(k)\, c_{k,\sigma}^{\dagger}\, c_{k,\sigma} + V \sum_{k,k'} c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger}\, c_{k'\downarrow}\, c_{-k'\uparrow} \tag{6.20}$$

6.3.3. the Mean field approximation:

In BEC, the order parameter is $\langle b^{+} \rangle$. Here, pairs of electrons forms a BEC state, so the order parameter is $\sum_k \langle c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger} \rangle$. In other words, in the condensed phase, $\langle c^{\dagger}\, c^{\dagger} \rangle \neq 0$. We can write:

$$c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger} = \langle c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger} \rangle + \left(c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger} - \langle c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger} \rangle \right) \tag{6.21}$$

$$c_{k'\downarrow}\, c_{-k'\uparrow} = \langle c_{k'\downarrow}\, c_{-k'\uparrow} \rangle + \left(c_{k'\downarrow}\, c_{-k'\uparrow} - \langle c_{k'\downarrow}\, c_{-k'\uparrow} \rangle \right) \tag{6.22}$$

The physical meaning for these two formulas are: the operator $c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger}$ has average value $\langle c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger} \rangle$. On top of the average value, the operator fluctuates around this average value with fluctuations being $c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger} - \langle c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger} \rangle$. If we believe that the fluctuations of the operator is small, the last term in the formula above is small. Here we can define $X_{k'} = \langle c_{k'\downarrow}\, c_{-k'\uparrow} \rangle$, as a result, $X_k^* = \langle c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger} \rangle$. So we have

$$c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger} = X_k^* + \left(c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger} - X_k^* \right) \tag{6.23}$$

$$c_{k'\downarrow}\, c_{-k'\uparrow} = X_{k'} + \left(c_{k'\downarrow}\, c_{-k'\uparrow} - X_{k'} \right) \tag{6.24}$$

We rewrite the interactions in terms of the average values $X_{k'}$ and X_k^* and fluctuations, $\left(c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger} - X_k^* \right)$ and $\left(c_{k'\downarrow}\, c_{-k'\uparrow} - X_{k'} \right)$

$$\frac{V}{2} \sum_{k} c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger}\, c_{k'\downarrow}\, c_{-k'\uparrow} = V \sum_{k,k'} \left[X_k^* + \left(c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger} - X_k^* \right) \right]\left[X_{k'} + \left(c_{k'\downarrow}\, c_{-k'\uparrow} - X_{k'} \right) \right] =$$
$$V \sum_{k,k'} \left[X_k^* X_{k'} + X_k^* \left(c_{k'\downarrow}\, c_{-k'\uparrow} - X_{k'} \right) + \left(c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger} - X_k^* \right) X_{k'} + \left(c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger} - X_k^* \right)\left(c_{k'\downarrow}\, c_{-k'\uparrow} - X_{k'} \right) \right]. \tag{6.25}$$

The last term is fluctuations times fluctuations. If fluctuations is small, the last term would be the smallest part.

Mean-field approximation: ignore the fluctuations*fluctuations part, so the Hamiltonian becomes:

$$H_{MF} = \sum_{k} \left[\epsilon(k)\, c_{k,\uparrow}^{\dagger}\, c_{k,\uparrow} + \epsilon(k)\, c_{k,\downarrow}^{\dagger}\, c_{k,\downarrow} \right] + V \sum_{k} X_k^* \sum_{k'} c_{k'\downarrow}\, c_{-k'\uparrow} + V \sum_{k} X_k \sum_{k} c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger} - V \sum_{k,k'} X_k^* X_{k'} \tag{6.26}$$

Within the mean-field approximation, the mean-field Hamiltonian only have quadratic terms in c and $c^{\dagger}$ (very similar to an free Fermi system).

Define $\Delta^* = V \sum_k X_k^*$ and $\Delta = V \sum_k X_k$,

$$H_{MF} = \sum_{k} \left[\epsilon(k)\, c_{k,\uparrow}^{\dagger}\, c_{k,\uparrow} + \epsilon(k)\, c_{k,\downarrow}^{\dagger}\, c_{k,\downarrow} + \Delta^* c_{k\downarrow}\, c_{-k\uparrow} + \Delta \sum_k c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger} \right] - \frac{\Delta^* \Delta}{V} \tag{6.27}$$

If we work with fixed chemical potential, μ, it make s more sense to consider $H - \mu N$, instead of H

$$H_{MF} - \mu N = \sum_{k} \left\{ [\epsilon(k) - \mu]\, c_{k,\uparrow}^{\dagger}\, c_{k,\uparrow} + [\epsilon(k) - \mu]\, c_{k,\downarrow}^{\dagger}\, c_{k,\downarrow} + \Delta^* c_{k\downarrow}\, c_{-k\uparrow} + \Delta\, c_{k\uparrow}^{\dagger}\, c_{-k\downarrow}^{\dagger} \right\} - \frac{\Delta^* \Delta}{V} \tag{6.28}$$

We can write it in terms of a two-by-two matrix

$$H_{\text{MF}} - \mu N = \sum_k \left(c_{k,\uparrow}{}^\dagger \quad c_{-k,\downarrow} \right) \begin{pmatrix} \epsilon(k) - \mu & \Delta \\ \Delta^* & \mu - \epsilon(k) \end{pmatrix} \begin{pmatrix} c_{k,\uparrow} \\ c_{-k,\downarrow}{}^\dagger \end{pmatrix} - \frac{\Delta^* \Delta}{V} \tag{6.29}$$

This is very similar to the Hamiltonian of a 2-band systems.

6.3.4. Bogoliubov transformation

Similar as in a 2-band systems, we should diagonalize the matrix using a unitary transformation, and here, this transformation is known as the Bogoliubov transformation. Define:

$$\gamma_{1k} = u\, c_{k,\uparrow} + v\, c_{-k,\downarrow}{}^\dagger \tag{6.30}$$

$$\gamma_{2k} = -v^*\, c_{k,\uparrow} + u^*\, c_{-k,\downarrow}{}^\dagger \tag{6.31}$$

where u and v are complex numbers with $|u|^2 + |v|^2 = 1$. It is easy to check that the conjugate operators are

$$\gamma_{1k}{}^\dagger = u^*\, c_{k,\uparrow}{}^\dagger + v^*\, c_{-k,\downarrow} \tag{6.32}$$

$$\gamma_{2k}{}^\dagger = -v\, c_{k,\uparrow}{}^\dagger + u\, c_{-k,\downarrow} \tag{6.33}$$

The inverse transformation is:

$$c_{k,\uparrow} = u^*\, \gamma_{1k} - v\, \gamma_{2k} \tag{6.34}$$

$$c_{-k,\downarrow}{}^\dagger = v^*\, \gamma_{1k} + u\, \gamma_{2k} \tag{6.35}$$

and

$$c_{k,\uparrow}{}^\dagger = u\, \gamma_{1k}{}^\dagger - v^*\, \gamma_{2k}{}^\dagger \tag{6.36}$$

$$c_{-k,\downarrow} = v\, \gamma_{1k}{}^\dagger + u^*\, \gamma_{2k}{}^\dagger \tag{6.37}$$

It is also easy to check that the new operators γ_{1k}, γ_{2k}, $\gamma_{1k}{}^\dagger$ and $\gamma_{2k}{}^\dagger$ satisfies the anti-commutation relations (so they are fermions). Here, I check on anti-commutation relation

$$\{\gamma_{1k}{}^\dagger, \gamma_{1k}\} =$$
$$\{u\, c_{k,\uparrow}{}^\dagger + v\, c_{-k,\downarrow}, u\, c_{k,\uparrow} + v\, c_{-k,\downarrow}{}^\dagger\} = u^2 \{c_{k,\uparrow}{}^\dagger, c_{k,\uparrow}\} + v^2 \{c_{-k,\downarrow}, c_{-k,\downarrow}{}^\dagger\} + u\,v \{c_{k,\uparrow}{}^\dagger, c_{-k,\downarrow}{}^\dagger\} + u\,v \{c_{-k,\downarrow}, c_{-k,\downarrow}\} = u^2 + v^2 = 1. \tag{6.38}$$

Therefore, these operators creates and annihilates fermions, and these fermions are known as Bogoliubov quasi-particles.

$$H_{\text{MF}} - \mu N = \sum_k \left(\gamma_{1k}{}^\dagger \quad \gamma_{2k}{}^\dagger \right) \cdot \begin{pmatrix} u & v \\ -v^* & u^* \end{pmatrix} \cdot \begin{pmatrix} \epsilon(k) - \mu & \Delta \\ \Delta^* & \mu - \epsilon(k) \end{pmatrix} \cdot \begin{pmatrix} u^* & -v \\ v^* & u \end{pmatrix} \begin{pmatrix} \gamma_{1k} \\ \gamma_{2k} \end{pmatrix} - \frac{\Delta^* \Delta}{V} \tag{6.39}$$

$$\begin{pmatrix} u & v \\ -v^* & u^* \end{pmatrix} \cdot \begin{pmatrix} \epsilon(k) - \mu & \Delta \\ \Delta^* & \mu - \epsilon(k) \end{pmatrix} \cdot \begin{pmatrix} u^* & -v \\ v^* & u \end{pmatrix} = \begin{pmatrix} u & v \\ -v^* & u^* \end{pmatrix} \cdot \begin{pmatrix} (\epsilon(k) - \mu)\, u^* + \Delta\, v^* & -(\epsilon(k) - \mu)\, v + \Delta\, u \\ \Delta^*\, u^* + (\mu - \epsilon(k))\, v^* & -\Delta^*\, v + (\mu - \epsilon(k))\, u \end{pmatrix} =$$
$$\begin{pmatrix} (\epsilon(k) - \mu)\, u^*\, u + \Delta\, v^*\, u + \Delta^*\, u^*\, v + (\mu - \epsilon(k))\, v^*\, v & -2\,(\epsilon(k) - \mu)\, v\, u + \Delta\, u^2 - \Delta^*\, v^2 \\ -2\,(\epsilon(k) - \mu)\, u^*\, v^* - \Delta\, v^{*2} + \Delta^*\, u^{*2} & (\epsilon(k) - \mu)\, v\, v^* - \Delta\, u\, v^* - \Delta^*\, u^*\, v + (\mu - \epsilon(k))\, u^*\, u \end{pmatrix} \tag{6.40}$$

Since we want to diagonalize the matrix, $-2\,(\epsilon(k) - \mu)\, v\, u + \Delta\, u^2 - \Delta^*\, v^2 = 0$

$$2\,(\epsilon(k) - \mu)\, v\, u = \Delta\, u^2 - \Delta^*\, v^2 \tag{6.41}$$

Define:

$$u = e^{i\phi_u} \cos \chi \quad \text{and} \quad v = e^{i\phi_v} \sin \chi \quad \text{and} \quad \Delta = |\Delta|\, e^{i\phi} \tag{6.42}$$

$$2\,(\epsilon(k) - \mu) \cos \chi \sin \chi\, e^{i\phi_u + i\phi_v} = e^{2i\phi_u + i\phi}\, \Delta\, \cos^2 \chi - e^{2i\phi_v - i\phi}\, \Delta^* \sin^2 \chi \tag{6.43}$$

So we have

$$\phi_u + \phi_v = 2\,\phi_u + \phi = 2\,\phi_v - \phi \tag{6.44}$$

and

$$2 (\epsilon(k) - \mu) \cos \chi \sin \chi = |\Delta| \left(\cos^2 \chi - \sin^2 \chi \right) \tag{6.45}$$

For the phases, we get $\phi_u - \phi_v = -\phi$, while $\phi_u + \phi_v$ can take any value. Without loss of generality, we can choose $\phi_u = -\phi_v = -\phi/2$. For χ we have,

$$(\epsilon(k) - \mu) \sin 2 \chi = |\Delta| \cos 2 \chi \tag{6.46}$$

So $\tan 2 \chi = \frac{|\Delta|}{\epsilon(k) - \mu}$.

In conclusion, if we choose $u = e^{-i \phi/2} \cos \chi$ and $v = e^{i \phi/2} \sin \chi$ with $\chi = \frac{1}{2} \arctan \frac{|\Delta|}{\epsilon(k) - \mu}$, the two-by-two matrix turns into a diagonal matrix:

$$\begin{pmatrix} (\epsilon(k) - \mu) \cos 2 \chi + \Delta \sin 2 \chi & 0 \\ 0 & -(\epsilon(k) - \mu) \cos 2 \chi - \Delta \sin 2 \chi \end{pmatrix} = \begin{pmatrix} \sqrt{(\epsilon(k) - \mu)^2 + |\Delta|^2} & 0 \\ 0 & -\sqrt{(\epsilon(k) - \mu)^2 + |\Delta|^2} \end{pmatrix} \tag{6.47}$$

$$H_{\text{MF}} - \mu N = \sum_k \left(\gamma_{1k}{}^\dagger \quad \gamma_{2k}{}^\dagger \right) \cdot \begin{pmatrix} E_k & 0 \\ 0 & -E_k \end{pmatrix} \begin{pmatrix} \gamma_{1k} \\ \gamma_{2k} \end{pmatrix} - \frac{\Delta^* \Delta}{V} \tag{6.48}$$

with $E_k = \sqrt{(\epsilon_k - \mu)^2 + (|\Delta|)^2}$.

For the top band, we have energy $E_k \geq |\Delta|$. for the bottom band, the dispersion relation $-E_k \leq -|\Delta|$. So we have energy gap 2|Δ|, if |Δ|>0.

A superconductor is an insulator of Bogoliubov quasi-particles!

6.3.5. Δ=?

The value of Δ is determined by self-consistent condition.

$$\Delta = V \sum_k \langle c_{-k\downarrow} c_{k\uparrow} \rangle = V \sum_k \langle (v \gamma_{1k}{}^\dagger + u^* \gamma_{2k}{}^\dagger)(u^* \gamma_{1k} - v \gamma_{2k}) \rangle =$$
$$V \sum_k \left[u^* v \langle \gamma_{1k}{}^\dagger \gamma_{1k} \rangle - u^* v \langle \gamma_{2k}{}^\dagger \gamma_{2k} \rangle \right] = V \sum_k \left[u^* v \frac{1}{\exp\left[\frac{E_k}{k_B T}\right] + 1} - u^* v \frac{1}{\exp\left[\frac{-E_k}{k_B T}\right] + 1} \right] \tag{6.49}$$

$$|\Delta| e^{i \phi} = V e^{i \phi} \sum_k \frac{\sin 2 \chi}{2} \left[\frac{1}{\exp\left[\frac{E_k}{k_B T}\right] + 1} - \frac{1}{\exp\left[\frac{-E_k}{k_B T}\right] + 1} \right] \tag{6.50}$$

$$|\Delta| = V \sum_k \frac{\sin 2 \chi}{2} \left[\frac{\exp\left[-\frac{E_k}{2 k_B T}\right]}{\exp\left[\frac{E_k}{2 k_B T}\right] + \exp\left[-\frac{E_k}{2 k_B T}\right]} - \frac{\exp\left[\frac{E_k}{2 k_B T}\right]}{\exp\left[-\frac{E_k}{2 k_B T}\right] + \exp\left[\frac{E_k}{2 k_B T}\right]} \right] =$$

$$V \sum_k \frac{\sin 2 \chi}{2} \frac{\exp\left[-\frac{E_k}{2 k_B T}\right] - \exp\left[\frac{E_k}{2 k_B T}\right]}{\exp\left[\frac{E_k}{2 k_B T}\right] + \exp\left[-\frac{E_k}{2 k_B T}\right]} = -V \sum_k \frac{|\Delta|}{2 \sqrt{(\epsilon(k) - \mu)^2 + |\Delta|^2}} \tanh\left(\frac{E_k}{2 k_B T}\right) \tag{6.51}$$

$$|\Delta| = -V \sum_k \frac{|\Delta|}{2 \sqrt{(\epsilon(k) - \mu)^2 + |\Delta|^2}} \tanh\left(\frac{E_k}{2 k_B T}\right) \tag{6.52}$$

|Δ|=0 is always a solution for this equation. If |Δ|≠0,

$$1 = -\frac{V}{2} \sum_k \frac{1}{\sqrt{(\epsilon(k) - \mu)^2 + |\Delta|^2}} \tanh\left(\frac{\sqrt{(\epsilon(k) - \mu)^2 + |\Delta|^2}}{2 k_B T}\right) \tag{6.53}$$

For repulsive interactions, this equation has no solution, because the l.h.s.>0 while the r.h.s.<0. However, for V<0, we have

$$1 = \frac{|V|}{2} \sum_k \frac{1}{\sqrt{(\epsilon(k) - \mu)^2 + |\Delta|^2}} \tanh\left(\frac{\sqrt{(\epsilon(k) - \mu)^2 + |\Delta|^2}}{2 k_B T}\right) \tag{6.54}$$

The equations may have some solution. In a Fermi liquid, typically only fermions near the Fermi surface will contribute and near the Fermi surface, the sum of momentum k can be written as an integer of energy.

$$1 = \frac{|V|}{2} \int N(0) \, d\epsilon \, \frac{1}{\sqrt{(\epsilon - \mu)^2 + |\Delta|^2}} \tanh\left(\frac{\sqrt{(\epsilon - \mu)^2 + |\Delta|^2}}{2 k_B T}\right) \tag{6.55}$$

where N(0) is known as the density of the states, which can be measure directly in STM. The range of the interaction is determined by the Debye frequency, which is the typical energy scales of the phonons (remember that the attractive interaction comes from phonons).

$$1 = \frac{|V|}{2} N(0) \int_{\mu - \epsilon_D}^{\mu + \epsilon_D} d\epsilon \, \frac{\tanh\left(\sqrt{(\epsilon(k) - \mu)^2 + |\Delta|^2}\,/2 k_B T\right)}{\sqrt{(\epsilon(k) - \mu)^2 + |\Delta|^2}} \tag{6.56}$$

If we define $\xi = \epsilon - \mu$,

$$1 = \frac{|V|}{2} N(0) \int_{-\epsilon_D}^{+\epsilon_D} d\xi \, \frac{\tanh\left(\sqrt{\xi^2 + |\Delta|^2}\,/2 k_B T\right)}{\sqrt{\xi^2 + |\Delta|^2}} = |V| N(0) \int_0^{+\epsilon_D} d\xi \, \frac{\tanh\left(\sqrt{\xi^2 + |\Delta|^2}\,/2 k_B T\right)}{\sqrt{\xi^2 + |\Delta|^2}} \tag{6.57}$$

For $T > T_c$, there is only one solution for the self-consistency equation, $\Delta = 0$.

For $T < T_c$, we have one solution $\Delta = 0$, which corresponds to the unstable solution for the G-L free energy. In addition, we also have a solution with $|\Delta| = \Delta(T)$, where $\Delta(T)$ is a function of T.

Because $|\Delta| = 0$ at T_c, we know that

$$1 = |V| N(0) \int_0^{+\epsilon_D} d\xi \, \frac{\tanh(\xi/2 k_B T_C)}{\xi} = |V| N(0) \int_0^{\epsilon_D/2 k_B T_C} dx \, \frac{\tanh(x)}{x} \tag{6.58}$$

In the weakly coupling limit [|V|<<1/N(0)], $k_B T_c \ll \epsilon_D$, so the upper limit of the integral is very large $\epsilon_D/2 k_B T_C \gg 1$. For this limit,

$$1 = |V| N(0) \int_0^{\epsilon_D/2 k_B T_C} dx \, \frac{\tanh(x)}{x} \approx N(0) |V| \ln\left(1.13 \frac{\epsilon_D}{k_B T_C}\right) \tag{6.59}$$

So,

$$T_c = 1.13 \frac{\epsilon_D}{k_B} \exp\left[-\frac{1}{N(0) |V|}\right] \tag{6.60}$$

The critical temperature is proportional to the Debye frequency. In addition, it depends on the density of states N(0) and interaction strengths |V|. As the interaction becomes weaker and weaker, T_c goes down to zero.
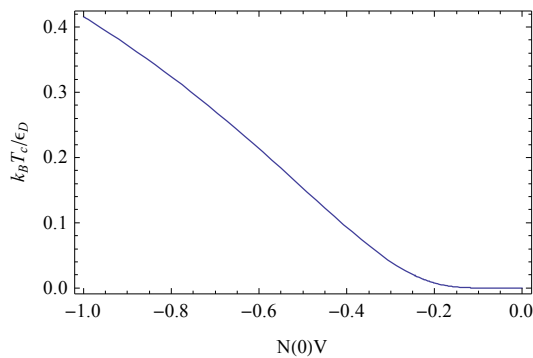
Fig. 5. T_c as a function of the interaciton strength.

Order parameter at zero temperature can be found as:

$$1 = \left| V \right| N(0) \int_0^{+\epsilon_D} d\xi \, \frac{1}{\sqrt{\xi^2 + |\Delta|^2}} = \left| V \right| N(0) \ln\left[\frac{\epsilon_D + \sqrt{|\Delta_0|^2 + \epsilon_D^2}}{|\Delta_0|} \right] \tag{6.61}$$

In the weak coupling limit [$V \, N(0) << 1$], we have $|\Delta_0| << \epsilon_D$

$$1 = \left| V \right| N(0) \ln\left[\frac{\epsilon_D + \sqrt{|\Delta_0|^2 + \epsilon_D^2}}{|\Delta_0|} \right] \approx \left| V \right| N(0) \ln \frac{2 \, \epsilon_D}{|\Delta_0|} \tag{6.62}$$

$$\Delta_0 = 2 \, \epsilon_D \exp\left(-\frac{1}{N(0) \, |V|} \right) \tag{6.63}$$

In particular, the ratio between Δ_0 and T_c is a universal constant:

$$\frac{\Delta_0}{k_B \, T} = \frac{2}{1.13} = 1.764 \tag{6.64}$$

This is true for all weakly-correlated superconductors.

$$1 = \left| V \right| N(0) \int_0^{+\epsilon_D} d\xi \frac{1}{\sqrt{\xi^2 + |\Delta|^2}} = \left| V \right| N(0) \ln\left[\frac{\epsilon_D + \sqrt{|\Delta_0|^2 + \epsilon_D^2}}{|\Delta_0|} \right] \tag{6.61}$$

In the weak coupling limit $[V\,N(0) << 1]$, we have $|\Delta_0| << \epsilon_D$

$$1 = \left| V \right| N(0) \ln\left[\frac{\epsilon_D + \sqrt{|\Delta_0|^2 + \epsilon_D^2}}{|\Delta_0|} \right] \approx \left| V \right| N(0) \ln \frac{2\,\epsilon_D}{|\Delta_0|} \tag{6.62}$$

$$\Delta_0 = 2\,\epsilon_D \exp\left(-\frac{1}{N(0)\,|V|} \right) \tag{6.63}$$

In particular, the ratio between Δ_0 and T_c is a universal constant:

$$\frac{\Delta_0}{k_B\,T} = \frac{2}{1.13} = 1.764 \tag{6.64}$$

This is true for all weakly-correlated superconductors.

6.4. Effective guage theory

6.4.1. Gauge symmetry

In a superconductor, the complex order parameter $\psi(r) = \langle c(r)\,c(r) \rangle$ is nonzero. By minimizing the GL free energy, we fixed the value of $|\psi(r)|$, but the phase of $\psi(r)$ is not determined.

Q: What will happen when we perform a gauge transformation?

The gauge field:

$$A_\mu(r) \rightarrow A_\mu(r) + \partial_\mu \Lambda(r) \tag{6.65}$$

The matter field:

$$c(r) \rightarrow c(r) \exp(i\,e\,\Lambda(r)/\hbar) \tag{6.66}$$

The order parameter:

$$\psi(r) = \langle c(r)\,c(r) \rangle \rightarrow \langle c(r) \exp(i\,e\,\Lambda(r)/\hbar)\,c(r) \exp(i\,e\,\Lambda(r)/\hbar) \rangle = \psi(r) \exp(2\,i\,e\,\Lambda(r)/\hbar) \tag{6.67}$$

The charge "2e" is because a cooper pair contains 2 electrons.

Derivative of $\psi(r)$

$$\partial_\mu \psi(r) \rightarrow \partial_\mu[\psi(r) \exp(2\,i\,e\,\Lambda(r)/\hbar)] = \partial_\mu \psi(r) + 2\,i\,e/\hbar\,\partial_\mu \Lambda(r)\,\psi(r) \tag{6.68}$$

Therefore, it is easy to check that $\left[\partial_\mu - 2\,i\,e\,A_\mu(r)/\hbar \right] \psi(r)$ is gauge invariant.

$$\partial_\mu \psi(r) - 2\,i\,e\,A_\mu(r)\,\psi(r)/\hbar \rightarrow \partial_\mu \psi(r) - 2\,i\,e\,A_\mu(r)\,\psi(r)/\hbar \tag{6.69}$$

This implies that if we write down the Lagrangian, L, any spatial derivative of the order parameter ($\partial_\mu \psi(r)$) must be combined with a term $-2\,i\,e\,A_\mu(r)\,/\,\hbar$. In fact, in order to keep the gauge symmetry, we need to replace any $\partial_\mu \psi$ by $\left[\partial_\mu - 2\,i\,e\,A_\mu(r)/\hbar \right] \psi(r)$. This is just the minimal coupling for charged particles we discussed before (replacing the momentum $p_\mu = -i\,\hbar\,\partial_\mu$ by $p_\mu + q\,A = -i\,\hbar\,\partial_\mu + q\,A_\mu$)

If we only focus on the phase of the order parameter, $\psi(r) = |\psi| \exp[i\,\phi(r)]$

$$\begin{aligned} \left[\partial_\mu - 2\,i\,e\,A_\mu(r)/\hbar \right] \psi(r) &= \left| \psi \right| \left[\partial_\mu - 2\,i\,e\,A_\mu(r)/\hbar \right] \exp[i\,\phi(r)] = \left| \psi \right| \exp[i\,\phi(r)] \left[i\,\partial_\mu \phi(r) - 2\,i\,e\,A_\mu(r)/\hbar \right] = \\ &\quad i\,\left| \psi \right| \exp[i\,\phi(r)] \left[\partial_\mu \phi(r) - 2\,e\,A_\mu(r)/\hbar \right] \propto \partial_\mu \phi(r) - 2\,e\,A_\mu(r)/\hbar \end{aligned} \tag{6.70}$$

6.4.2. the Lagrangian.

If we write down the Lagrangian for a superconductor, we know that it must contain gauge fields and the phase of the order parameter. The electrons should not appear if we only care about low enegy physics smaller than the superconducting gap (remember that supercondcutors are

electrons should not appear if we only care about low enegy physics smaller than the superconducting gap (remember that supercondcutors are insualtors for Bogoliubov quasi-particles, so the fermions cannot move similar to an insulator. Similar to an insulator, we don't need to consider the motion of electrons). The amplitude of the order parameter is not important either, because their value is fixed by the GL free energy. So have can write down the theory as:

$$L = \frac{1}{2} \int d\vec{r} \left(\epsilon_0 E^2 - \frac{B^2}{\mu_0} \right) + L_S\left[\partial_\mu \phi(r) - 2 e \, A_\mu(r)\big/\hbar \right] \tag{6.71}$$

It is important to notice that the superconductor part L_S cannot depends on ϕ itself, because ϕ is not gauge invariant. However, it can be a function of $\partial_\mu \phi(r) - 2 e \, A_\mu(r)\big/\hbar$, which is gauge invariant.

for simplicity, lets use the theorist units: $2e=\hbar=1$.

$$L = \frac{1}{2} \int d\vec{r} \left(\epsilon_0 E^2 - \frac{B^2}{\mu_0} \right) + L_S\left[\partial_\mu \phi(r) - A_\mu(r) \right] \tag{6.72}$$

6.4.3.charge and currents

Currents and charge

$$J_i = \frac{\delta L_S}{\delta A_i} = -\frac{\delta L_S}{\delta \, \partial_i \phi} \tag{6.73}$$

$$\rho = \frac{\delta L_S}{\delta A_0} = -\frac{\delta L_S}{\delta \, \partial_0 \phi} \tag{6.74}$$

6.4.4.Hamiltonian and $\partial_\mu \phi = A_\mu$

$$H = \frac{1}{2} \int d\vec{r} \left(E^2 + B^2 \right) + H_S\left[\partial_\mu \phi(r) - A_\mu(r) \right] \tag{6.75}$$

If we do a Taylor expansion,

$$H = \frac{1}{2} \int d\vec{r} \left(E^2 + B^2 \right) + \frac{\kappa_0}{2} \int d\vec{r} \left[\partial_0 \phi(r) - A_0(r) \right]^2 + \sum_i \frac{\kappa_i}{2} \int d\vec{r} \left[\partial_i \phi(r) - A_i(r) \right]^2 + \text{higher order terms} \tag{6.76}$$

κ_0 and κ_i are positive due to stability reasons.

To minimize the energy, $\partial_\mu \phi = A_\mu$. This equation is the key to understand superconductivity. It means that in a superconductor, the gauge field and the phase of the order parameters are directly connected to each other. This relation is a unique property of superconductors.

A guage field which is a total derivative is known as a "pure gauge'. It means that if the system is singly-connected, there is no E=B=0. In a singly connected space, we can define a gauge transformation $A_\mu \to A_\mu + \partial_\mu \Lambda$ with $\Lambda = -\phi$. After this gauge transformation, $A'_\mu = 0$. so E=B=0.

6.4.5. Meissner effect and magnetic levitation

Because $\partial_\mu \phi = A_\mu$, the magnetic field

$$B = \nabla \times A = \nabla \times \nabla \phi = 0 \tag{6.77}$$

There is no magnetic field inside a superconductor. This is the Meissner effect.

6.4.6. Critical field B_c

The Meissner effect expels the B field, which costs energy $B^2/2$ (times the size of the system). If this energy cost is larger than the energy gain to form Cooper pairs Δ, the system will not want to form a superconducting state. This defines a critical field $B_C = \sqrt{2\Delta}$. If B field is above this value, the system turns into a normal metal (paramagnetic)

6.4.7. quantization of a magnetic flux.

If we have a superconducting ring, we can integrate the gauge field around the ring, which is just the magnetic flux.

$$\oint A \cdot dr = \int\int B \cdot dS \tag{6.78}$$

We know that $\partial_\mu \phi = A_\mu$, so

$$\oint A \cdot dr = \oint \partial_\mu \phi \cdot dr = 2 n \pi \tag{6.79}$$

Here we used the fact that the order parameter is single-valued. Therefore, if we go around the ring, the phase of the order parameter can only change by $2 n \pi$. If we put the unit back,

$$\oint A \cdot dr = \oint \partial_\mu \phi \cdot dr = 2 n \pi \frac{\hbar}{2 e} = \frac{n \pi \hbar}{e} \tag{6.80}$$

Notice that the denominator here is 2e instead of e. This means that the fundamental particles in a superconductor carries electronic charge 2e, which is a cooper pair. The quantization of the magnetic flux (to $2\pi\hbar/2e$ instead of $2\pi \hbar/e$) is a direct experimental evidence for the formation of Cooper pairs.

This is also one of the best techniques to measure $\hbar$

6.4.8. Zero resistivity

We know

$$\rho = \frac{\delta L_S}{\delta A_0} = -\frac{\delta L_S}{\delta \partial_0 \phi} \tag{6.81}$$

We also know that the canonical momentum π is defined as

$$\pi = \frac{\delta L_S}{\delta \partial_0 \phi} \tag{6.82}$$

So, $\rho = -\pi$. Therefore, the Hamiltonian should be considered a function of $-\rho$ and ϕ, instead of $\partial_0 \phi$ and ϕ.

The equation of motion is:

$$\partial_0 \phi = \frac{\delta H_S}{\delta \pi} = -\frac{\delta H_S}{\delta \rho} \tag{6.83}$$

On the other hand, we know that

$$\frac{\delta H_S}{\delta \rho(r)} = V(r) \tag{6.84}$$

where V is the voltage. This is because $E = \int \rho(r) V(r) dr$

Therefore,

$$\partial_0 \phi(r) = -V(r) \tag{6.85}$$

For a static system, the phase of the order parameter should be time-independent, $\partial_0 \phi = 0$. So we have $V(r)=0$.

This means that if we have a constant current in the system (the system is static), $V = 0$. No voltage but finite current. This is superconductivity.

6.4.9. The Josephson junction.

Josephson junction has two superconductors separated by a thing layer of insulator. The Lagrange of a Josephson junction L_J depends on the phase difference between the two superconductors, which is also gauge invariant

$$L = L_{E \text{ and } B} + L_{S1} + L_{S2} + L_{\text{junction}} \tag{6.86}$$

$$L_{\text{junction}} = \mathcal{A} F(\Delta\phi) \tag{6.87}$$

where $\mathcal{A}$ is the area of the junction. $F(\Delta\phi)$ is a function of the phase difference between the two superconductors ($\Delta\phi = \phi_L - \phi_R$).

It is easy to notice that $F(\Delta\phi)$ must be a periodic function with periodicity $2 n \pi \hbar/2 e = n \pi \hbar/e$. Again the denominator 2 e is due to the fact that Cooper pairs have charge 2e.

$$F(\Delta\phi) = F(\Delta\phi + n \pi \hbar/e) \tag{6.88}$$

In the presence of a gauge field A, the gauge invariance $\Delta\phi$ is

$$\Delta\phi = \int dr \cdot (\nabla\phi - A) \tag{6.89}$$

Therefore the current cross the junction is

$$J = \frac{\delta L_{\text{junction}}}{\delta A} = \mathcal{A} F'(\Delta\phi) \frac{\delta\Delta\phi}{\delta A} = \mathcal{A} F'(\Delta\phi) \frac{\delta A}{\delta A} = \mathcal{A} F'(\Delta\phi) \tag{6.90}$$

Now we apply a voltage V across the junction. Because we know that

$$\partial_0 \phi(r) = -V(r), \tag{6.91}$$

it is easy to notice that

$$\Delta\phi = -V t + \text{constant} \tag{6.92}$$

Therefore:

$$J = \mathcal{A} F'(-V t + \text{constant}) \tag{6.93}$$

By applying a fixed V, we found that the current is changing with time. Because F is a periodic function, F' is also a periodic function. So J is aperiodic function of t, and the periodicity is $\pi \hbar/e \, \Delta V$

$$\Delta\phi = \int dr \cdot (\nabla\phi - A) \tag{6.89}$$

Therefore the current cross the junction is

$$J = \frac{\delta L_{junction}}{\delta A} = \mathcal{A} F'(\Delta\phi) \frac{\delta\Delta\phi}{\delta A} = \mathcal{A} F'(\Delta\phi) \frac{\delta A}{\delta A} = \mathcal{A} F'(\Delta\phi) \tag{6.90}$$

Now we apply a voltage V across the junction. Because we know that

$$\partial_0 \phi(r) = -V(r), \tag{6.91}$$

it is easy to notice that

$$\Delta\phi = -V t + \text{constant} \tag{6.92}$$

Therefore:

$$J = \mathcal{A} F'(-V t + \text{constant}) \tag{6.93}$$

By applying a fixed V, we found that the current is changing with time. Because F is a periodic function, F' is also a periodic function. So J is aperiodic function of t, and the periodicity is $\pi \hbar/e \Delta V$

6.5. High Temperature superconductivity

High Tc compounds is a family of materials: La2-xSrxCuO4, Bi2Sr2Can-1CunO2n+4+x (BSCCO), YBa2Cu3O7-x, etc

6.5.1. Lattice structure

- Common feature: CuO2 planes separated by some barrier
- a undoped compound:
  - 1 hole per unit cell.
  - Cu site has lower potential energy, so electrons want to say on Copper
  - Strong Coulomb repulsion on the Cu site (it cost a lot of energy to put two holes on the same site). Basically, we can only have one hole per Cu site,
- The system is an insulator (Every Cu has one hole, then the hole cannot move any more).
- Each Cu site has one hole, and the hole carrier spin 1/2. So we have a lattice of spin. Anti-Ferromagnetic ordering.

6.5.2. Dope the system

Q: How to change electron density?

Option #1: Dope the system (replace some of the atom by another type of atom with different number of electrons, or remove some atoms from the systems).

- Dope electron or hole? Depending on the elements, we may be adding electron or removing electron from the system. Adding electron is know as "electron doping". Add hole is known as "hole doping".

Option #2, Gating

- too weak for high Tc compounds

Q: Which technique is good?

A: If one can choose from these two options, gating is in principle better, because it doesn't induce disorders. But in many cases gating only change the electron density by a small amount, so we have no choice by using the option #1.

Some high Tc compounds are hole doped and some are electron doped. But in most cases, "high Tc" means hole doped high Tc compounds, because hole doped case have a higher Tc.

6.5.3. Phase diagram
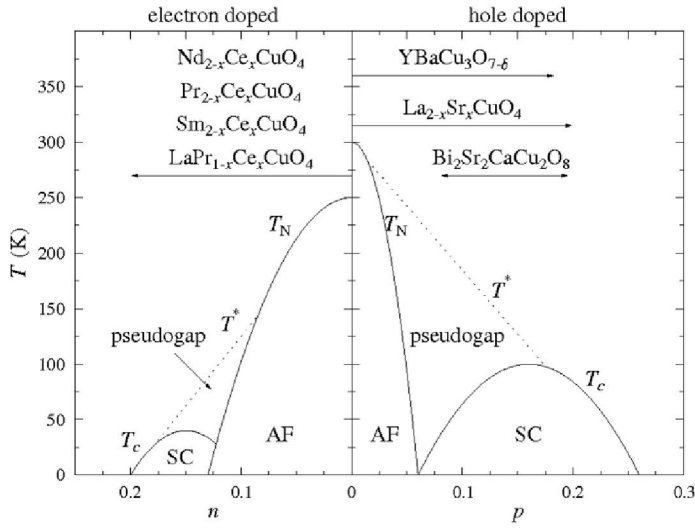
Fig. 6. Schematic Phase diagrams of hole-doped (right side) and electron doped (left side) high Tc superconductors (from wikipedia)

6.5.4. Tc
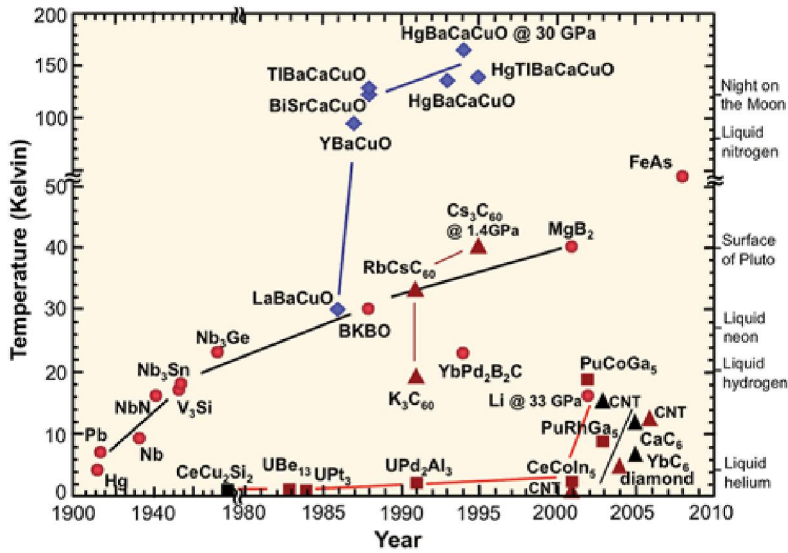
Fig. 7. Tc vs year

Iron based superconductor shows some similarity to electron doped high-Tc compounds (but they also have major differences). They have similar Tc. Both of them have strong e-e interaction, but the e-e interaction is less strong compare to hole-doped high Tc.

6.5.5. The order parameter (d-wave)

Order parameter is a function of momentum

$$\Delta_k = \langle c_{k\uparrow} c_{-k\downarrow} \rangle \tag{6.94}$$

In conventional superconductors (BCS theory), we assume Δ is isotropic, $\Delta_k = \langle c_{k\uparrow} c_{-k\downarrow} \rangle$ = constant. But in general it is a function of $\vec{k}$

$$\Delta_k = \Delta\left(\vec{k}\right) = \Delta(\,|\,k\,|,\,\theta_k) \tag{6.95}$$

The dependence on |k| is not important, because we know only electrons near the Fermi surface matters.

$$\Delta_k = \Delta\left(\vec{k}\right) = \Delta(\,|\,k\,|,\,\theta_k) \approx \Delta(k_F,\,\theta_k) = \Delta(\theta_k) \tag{6.96}$$

Let's consider the case $\Delta(\theta_k)$ is a real function. For conventional SCs and high Tc SCs, we can always make Δ real by choosing a proper gauge, but there are cases where Δ is NOT real and cannot be made real.

For conventional superconductors, $\Delta(\theta_k)$ varies as a function of θ, but the sign never change (s-wave)

For high T_c superconductors, $\Delta(\theta_k)$ changes sign for four times as we go around from $\theta_k = 0$ to $2\,\pi$. It has a d-wave symmetry

The superconducting phase of high Tc materials are described very well by a d-wave pairing state, if one don't ask why the electrons pair up at such a high temperature. The mystery of high Tc is in fact not about the superconducting phase, but about the normal phase. The normal phase is a bad metal and we don't really have a theoretical understand about it. Since we don't know the normal phase, we don't know why electron turns into a superconductor.

6.5.6. Normal phase

- Like a Fermi liquid at high doping.
- Not a Fermi liquid at low doping (known as nonFermi liquid or strange metal, or pseudo-gap phase)

Pseudo gap,

For the under-doped side, there is an temperature scale, T^*, at which the behavior of many experiential measurable quantity changes.

Different experiments give different T^*.

Below T^*, the density of state shows that there seems to be a insulating gap, but not a full insulating gap (a pseudo-gap).

Q: Whether this is related with the superconducting gap or this is due to some other ordering tendency?

A: Unclear.

Q: Whether T^* is a phase transition or a crossover?

A: Unclear. Because different experiments give different T^*, it is probably a crossover (phase transition should have a well-define transition temperature T_c and all experiment should give the same T_c). However, in the presence of disorder, the transition temperature T_c many have some distribution. So it is NOT impossible to think T^* as a phase transition.

6.5.7. Mechanics for superconductivity: under debating

What we know:

- Not phonon driven
  - too weak to get such a high Tc.
  - Isotope effect is weak.
  - Phonon prefer s-wave pairing
- Very likely, it is due to e-e interaction and probably involves spin
  - e-e interaction is very strong
  - spin-spin coupling is very strong
  - Tc is too high to be explained by other interactions, which are all too weak.

Q: How can repulsion lead to pairing?

A: Unclear, but we have some hint.

6.5.8. Attraction: superconductor vs. inhomogeneity

7

Frustrated magnets and 1D quantum system

7.1. magnetization and frustrated magnets (in any dimensions)

Consider a lattice of spins with spin interactions. The Heisenberg model:

$$H = J \sum_{<ij>} \vec{S}_i \cdot \vec{S}_j = J \sum_{<ij>} \left(S_i^x S_j^x + S_i^y S_j^y + S_i^z S_j^z \right) \tag{7.1}$$

If J<0, the spins want to point in the same direction as their neighbors (ferromagnetic). If J>0, neighboring spins want to point in the opposite directions to minimize the energy (anti-ferromagnetic).

In real materials (which may be anisotropic), the couplings in different directions may be different, which can be described by the XYZ model:

$$H = \sum_{<ij>} \left(J_x S_i^x S_j^x + J_y S_i^y S_j^y + J_z S_i^z S_j^z \right) \tag{7.2}$$

If $J_x = J_y$, we get the so called XXZ model

$$H = \sum_{<ij>} \left(J_x S_i^x S_j^x + J_x S_i^y S_j^y + J_z S_i^z S_j^z \right) \tag{7.3}$$

If $J_x = J_y$ and $J_z = 0$, we get the so called XY model

$$H = J \sum_{<ij>} \left(S_i^x S_j^x + S_i^y S_j^y \right) \tag{7.4}$$

If $J_x = J_y = 0$ and $J_z \neq 0$, we get the so called Ising model

$$H = J \sum_{<ij>} S_i^z S_j^z \tag{7.5}$$

7.1.1. Frustration

For ferromagnetic couplings, the ground state is simple: all spin pointing in the same directions (plus quantum fluctuations). For anti-ferromagnetic couplings, it depends on geometry. For example, on a square lattice, the spins can find a way to make all spins happy (all neighboring spins stay in opposite directions). But on a triangle, there is no way to make all the spin happy. So the spins are "frustrated", and they don't know in which direction they should be pointing. This type of systems are known as frustrated magnets.

The reason that frustrated magnets are interesting is because their ground states are not obvious. Therefore, some of these system may have very strange ground state, with exotic properties, such as spin liquids.

As T goes down, in most systems, some form of ordering will be formed (liquid turns in to solids, paramagnetism turns into ferromagnetism, etc.). Motion of particles, spins etc. are frozen at low T. However, if the quantum fluctuation is too strong, the system cannot order even at T=0 (liquid down to zero T). In spin systems, such a liquid is known as a spin liquid.

Typically, in theoretical studies, one maps spins into a strongly correlated Fermi gas (or boson gas), which is known as the slave fermion (or slave boson) approach. For the slave fermion approach, the fermion may form a insulator or superconductor (gaped spin liquid), semi-metal like graphene (gapless spin liquid, algebra spin liquid), or a metal (Bose metal and the Fermi surfaces for this metal is known as the spinon Fermi surface).

The mapping from spin to fermions/bosons can be rigorously constructed in any dimensions. However, one needs to pay a prize; these fermions/-

bosons have strong interactions, and thus the problem cannot be solved in general. However, in 1D, there is a very straightforward mapping to fermions and the fermions may not have any interactions (for the XY model). In 1D, bosons and fermions are not fundamentally different from each other.

7.2. Spin-1/2 and hard-core bosons (in any dimensions)

This section describes a baby version of the slave boson approach. Hard-core bosons are bosons hopping on a lattice, but we assume that there can be at most one boson per site. This physics can be realized if one introduces strong (short-range) repulsive interactions between bosons, so the boson will not want to stay together.

Consider one single lattice site, the size (dimension) of the Hilbert space of hard core bosons is 2: occupied or empty. For spin-1/2, the Hilbert space has the same size: spin up and spin down. So it is possible to make a connection between spins and hard-core bosons.

We can identify the spin up state for spin-1/2 with the occupied state for a hard core boson. Spin down state with an empty state.

Define hard core boson creation and annihilation operators b_i and $b_i^\dagger$.

$$b^\dagger |0\rangle = |1\rangle \qquad b|0\rangle = 0 \tag{7.6}$$

$$b^\dagger |1\rangle = 0 \qquad b|1\rangle = |0\rangle \tag{7.7}$$

Define spin rising and lowering operator $S_i^+ = S_x + i\, S_y$ and $S_i^- = S_x - i\, S_y$.

$$S^+ |\text{down}\rangle = |\text{up}\rangle \qquad S_i^- |\text{down}\rangle = 0 \tag{7.8}$$

$$S^+ |\text{up}\rangle = 0 \qquad S_i^- |\text{up}\rangle = |\text{down}\rangle \tag{7.9}$$

So, we find that $b^\dagger$ is S^+ and b is just S^-.

The density operator: $n = b^\dagger b$. It is easy to check that $n|1\rangle = 1 \times |1\rangle$ and $n|0\rangle = 0 \times |0\rangle$

For spin-1/2, the density operator turns into $n = S^+ S^- = (S_x + i\, S_y)(S_x - i\, S_y) = S_z + 1/2$.

$$n = S^+ S^- = (S_x + i\, S_y)(S_x - i\, S_y) =$$

$$S_x^2 + S_y^2 - i(S_x S_y - S_y S_x) = \left(\left|\vec{S}\right|\right)^2 - S_z^2 - i[S_x, S_y] = s(s+1) - S_z^2 + S_z = \frac{1}{2}\left(\frac{1}{2} + 1\right) - \frac{1}{4} + S_z = S_z + 1/2 \tag{7.10}$$

It is easy to check that $(S_z + 1/2)|\text{up}\rangle = 1 \times |\text{up}\rangle$ and $(S_z + 1/2)|\text{down}\rangle = 0 \times |\text{down}\rangle$

One can further verify the commutation relation.

$$[b, b] = [b^\dagger, b^\dagger] = 0 \tag{7.11}$$

$$[b, b^\dagger]|0\rangle = b\, b^\dagger |0\rangle - b^\dagger b |0\rangle = |0\rangle \tag{7.12}$$

$$[b, b^\dagger]|1\rangle = b\, b^\dagger |0\rangle - b^\dagger b |0\rangle = -|0\rangle \tag{7.13}$$

So,

$$[b, b^\dagger] = (1 - 2n) \tag{7.14}$$

For a lattice with different sites,

$$[b_i, b_j^\dagger] = (1 - 2n)\,\delta_{ij} \tag{7.15}$$

For spins, we have the same commutation relations.

$$[S_i^-, S_j^-] = [S_i^+, S_j^+] = 0 \tag{7.16}$$

$$[S_i^+, S_j^-] = [S_x + i\, S_y,\ S_x - i\, S_y] = -i[S_x, S_y] + i[S_y, S_x] = 2i[S_y, S_x] = 2 S_z = (2n - 1) = [b_j^\dagger, b_i] \tag{7.17}$$

7.2.1. XXZ model

$$H = \sum_{<ij>} J_x \left(S_i^x S_j^x + S_i^y S_j^y \right) + J_z S_i^z S_j^z =$$
$$\sum_{<ij>} J_x \left[S^+ S^- + i(S_x S_y - S_y S_x) \right] + J_z S_i^z S_j^z = \sum_{<ij>} J_x \left[(S^+ S^- - S_z) + J_z S_i^z S_j^z \right. \tag{7.18}$$

$$S_i^+ S_j^- + S_j^+ S_i^- = (S_i^x + i\, S_i^y)\left(S_j^x - i\, S_j^y\right) + \left(S_j^x + i\, S_j^y\right)\left(S_i^x - i\, S_i^y\right) =$$
$$2\, S_i^x S_j^x + 2\, S_i^y S_j^y + i\, S_i^y S_j^x - i\, S_i^x S_j^y + i\, S_j^y S_i^x - i\, S_j^x S_i^y = 2\, S_i^x S_j^x + 2\, S_i^y S_j^y \tag{7.19}$$

$$H = \frac{J_x}{2} \sum_{<ij>} \left(S_i^+ S_j^- + S_j^+ S_i^- \right) + J_z \sum_{<ij>} (n_i - 1/2)\left(n_j - 1/2\right) =$$
$$\frac{J_x}{2} \sum_{<ij>} \left(b_i^\dagger b_j + b_j^\dagger b_i \right) + J_z \sum_{<ij>} n_i n_j - J_z \sum_{(ij)} (n_i - 1/4) = \frac{J_x}{2} \sum_{<ij>} \left(b_i^+ b_j^- + b_j^+ b_i^- \right) + J_z \sum_{<ij>} n_i n_j - \text{constant} \tag{7.20}$$

$$H = t \sum_{<ij>} \left(b_i^\dagger b_j + b_j^\dagger b_i \right) + V \sum_{<ij>} n_i n_j \tag{7.21}$$

with $t = J_x/2$ and $V = J_z$.

It is worthwhile to emphasize that these bosons are hard core bosons, which are bosons with a huge on-site repulsion. Using ordinary bosons, this model becomes

$$H = t \sum_{<ij>} \left(a_i^\dagger a_j + a_j^\dagger a_i \right) + V \sum_{<ij>} n_i n_j + U \sum_i n_i(n_i - 1) \tag{7.22}$$

with $U \to \infty$. So this is a boson model with infinite interactions, which cannot be solved.

7.3. Spin-1/2 and fermions (1D)

This section is a baby version of the slave fermion approach. The problem of bosons is that we need infinite interactions to keep avoid many bosons occupying the same state. To make sure that the size of the Hilbert space matches, we need to ensure that there are at most 1 boson per site. If we use fermions, the problem of multi-occupation will not appear at all, due to the Pauli exclusive principle. So fermion seems to be our friends here. (Is it true?)

7.3.1. a single site

Consider one single lattice site, the size (dimension) of the Hilbert space of spinless fermion is 2 occupied or empty. For spin-1/2, the Hilbert space has the same size: spin up and spin down. So it seems possible to make a connection between spins and hard-core bosons.

We can identify the spin up state for spin-1/2 with the occupied state for a hard core boson. Spin down state with an empty state.

Define hard core boson creation and annihilation operators c_i and $c_i^\dagger$.

$$c^\dagger \left| 0 \right\rangle = \left| 1 \right\rangle \qquad\qquad c \left| 0 \right\rangle = 0 \tag{7.23}$$

$$c^\dagger \left| 1 \right\rangle = 0 \qquad\qquad c \left| 1 \right\rangle = \left| 0 \right\rangle \tag{7.24}$$

Define spin rising and lowering operator $S_i^+ = S_x + i\, S_y$ and $S_i^- = S_x - i\, S_y$.

$$S^+ \left| \text{down} \right\rangle = \left| \text{up} \right\rangle \qquad\qquad S_i^- \left| \text{down} \right\rangle = 0 \tag{7.25}$$

$$S^+ \left| \text{up} \right\rangle = 0 \qquad\qquad S_i^- \left| \text{up} \right\rangle = \left| \text{down} \right\rangle \tag{7.26}$$

So, we find that $c^\dagger$ is S^+ and c is just S^-.

The density operator: $n = c^\dagger c$. It is easy to check that $n \left| 1 \right\rangle = 1 \times \left| 1 \right\rangle$ and $n \left| 0 \right\rangle = 0 \times \left| 0 \right\rangle$

For spin-1/2, the density operator turns into $n = S^+ S^- = \left(S_x + i\, S_y\right)\left(S_x - i\, S_y\right) = S_z + 1/2$.

$$n = S^+ S^- = \left(S_x + i\, S_y\right)\left(S_x - i\, S_y\right) =$$
$$S_x^2 + S_y^2 - i\left(S_x S_y - S_y S_x\right) = \left(\left| \vec{S} \right| \right)^2 - S_z^2 - i\left[S_x, S_y\right] = s(s+1) - S_z^2 + S_z = \frac{1}{2}\left(\frac{1}{2} + 1\right) - \frac{1}{4} + S_z = S_z + 1/2 \tag{7.27}$$

It is easy to check that $(S_z + 1/2) | \text{up} \rangle = 1 \times | \text{up} \rangle$ and $(S_z + 1/2) | \text{down} \rangle = 0 \times | \text{down} \rangle$

One can further verify the anti-commutation relation.

$$\{c, c\} = \{c^\dagger, c^\dagger\} = 0 \tag{7.28}$$

$$\{c, c^\dagger\} = 1 \tag{7.29}$$

$$\{S^-, S^-\} = \{S^+, S^+\} = 0 \tag{7.30}$$

$$\{S^+, S^-\} = \{S_x + i S_y, \ S_x - i S_y\} = $$

$$2\left(S_x^2 + S_y^2\right) - i \ S_x S_y + i \ S_y S_x - i \ S_y S_x + i \ S_x S_y = 2\left(S_x^2 + S_y^2\right) = 2[s(s+1) - S_z^2] = 2\left[\frac{1}{2}\left(\frac{1}{2} + 1\right) - \frac{1}{4}\right] = 1 \tag{7.31}$$

7.3.2. two sites

$$\{c_i, c_j\} = \{c_i^\dagger, c_j^\dagger\} = 0 \tag{7.32}$$

$$\{S_i^-, S_j^-\} = \{S_i^+, S_j^+\} = 0 \tag{7.33}$$

$$\{c_1, c_2^\dagger\} = 0 \tag{7.34}$$

$$\{S_1^-, S_2^+\} \neq 0 \quad \text{but} \ [S_1^-, S_2^+] = 0 \tag{7.35}$$

One cannot naively identify S^+, S^- with $c^\dagger$ and c.

The correct mapping is

$$S_1^+ = c_1^\dagger \tag{7.36}$$

$$S_2^+ = c_2^\dagger (-1)^{n_1} \tag{7.37}$$

This extra factor $(-1)^{n_1}$ will generate the correct anti-commutation relation.

7.3.3. Many-sites

Consider many sites i=1,2,3…N

$$S_i^+ = c_i^\dagger (-1)^{\sum_{j=1}^{j=i-1} n_j} \tag{7.38}$$

This is known as the Jordan–Wigner transformation.

7.3.4. XXZ model

Mapping to fermions, we get

$$H = t \sum_{\langle ij \rangle} \left(c_i^\dagger c_j + c_j^\dagger c_i\right) + V \sum_{\langle ij \rangle} n_i n_j \tag{7.39}$$

with $t = J_x/2$ and $V = J_z$.

For XY model with $J_z = 0$, this is just a free fermion problem.

Notice that 1D is very special. In higher dimensions, the JW transformation will generate long range interactions. But in 1D there is no such problem.

8

Quantum field theory and Green's function approach

8.1. Background

8.1.1. Why QFT?

Same as second quantization. As long as one uses second quantization c and $c^\dagger$, QFT will emerge naturally.

Why second quantization? We want to handle large numbers of indistinguishable particles.

n indistinguishable particles:

$$\Psi(r_1, r_2, r_3, ..., r_n) = \sum_{\mathcal{P}} (\pm 1)^{\mathcal{P}} \, \psi_{i_1}(r_1) \, \psi_{i_2}(r_2) \, ... \psi_{i_n}(r_n) \tag{8.1}$$

where $\mathcal{P}$ represents all permutations and there are n! terms here. For large n, this is an extremely complicated wavefunction. For even ten particles, n=10, there are 2.6 million terms.

Same is true for high energy. Although the colliders focus on the collisions between two particles, there are many "virtue particles".

8.1.2. Expectation value at finite T (quantum statistical physics)

For classical statistical physics, we know that the expectation value of some quantity is (grand canonical ensemble)

$$\langle X \rangle = \frac{\sum_n X_n \exp[-\beta(E_n - \mu\, N)]}{\sum_n \exp[-\beta(E_n - \mu\, N)]} \tag{8.2}$$

For a quantum system, the expectation value is

$$\langle X \rangle = \frac{\sum_n \langle n \mid \hat{X} \exp[-\beta(\hat{H} - \mu\, \hat{N})] \mid n \rangle}{\sum_n \langle n \mid \exp[-\beta(\hat{H} - \mu\, \hat{N})] \mid n \rangle} = \frac{\mathrm{Tr}\left\{\hat{X} \exp[-\beta(\hat{H} - \mu\, \hat{N})]\right\}}{\mathrm{Tr}\left\{\exp[-\beta(\hat{H} - \mu\, \hat{N})]\right\}} \tag{8.3}$$

Here the $\sum_n$ is summing over a complete basis of the Hilbert space.

8.1.3. Second quantization

Operators: Every operator is written in terms of creation and annihilation operators c s and $c^\dagger$s: $\hat{X} = c_{k1}\, c_{k2}\, ... \, c_{kn}\, c_{q1}{}^\dagger\, c_{q2}{}^\dagger\, ...c_{qm}{}^\dagger$

Physical observables: Every observables are the expectation values of operators, which are products of cs and $c^\dagger$s: $\langle \hat{X} \rangle = \langle c_{k1}\, c_{k2}\, ... \, c_{kn}\, c_{q1}{}^\dagger\, c_{q2}{}^\dagger\, ...c_{qm}{}^\dagger \rangle$.

Tasks: compute these type of things: $\langle c_{k1}\, c_{k2}\, ... \, c_{kn}\, c_{q1}{}^\dagger\, c_{q2}{}^\dagger\, ...c_{qm}{}^\dagger \rangle$, which are known as correlation functions.

Consider a system with particle number conservation law (no superfluid, no superconductivity), it is easy to notice that $\langle c \rangle = \langle c^\dagger \rangle = 0$. This is because $\langle n|c|n \rangle = 0$.

In other words, the simplest correlation function one can imagine is $\langle c_k\, c_k{}^\dagger \rangle$

8.1.4. Normal order

Let's consider a system with Hamiltonian:

$$H = \int dr \, \frac{\nabla \psi^\dagger(r) \, \nabla \psi(r)}{2 \, m} + \frac{1}{2} \int dr \, dr' \, V(\,|\,r - r'\,|\,) \, \psi^\dagger(r) \, \psi(r) \, \psi^\dagger(r') \, \psi(r') =$$

$$\int dr \, \frac{\nabla \psi^\dagger(r) \, \nabla \psi(r)}{2 \, m} + \frac{1}{2} \int dr \, dr' \, V(\,|\,r - r'\,|\,) \, \psi^\dagger(r) \, \psi^\dagger(r') \, \psi(r') \, \psi(r) + \frac{1}{2} \int dr \, dr' \, V(\,|\,r - r'\,|\,) \, \delta(r - r') \, \psi^\dagger(r) \, \psi(r')$$

$$= \int dr \, \frac{\nabla \psi^\dagger(r) \, \nabla \psi(r)}{2 \, m} + \frac{1}{2} \int dr \, dr' \, V(\,|\,r - r'\,|\,) \, \psi^\dagger(r) \, \psi^\dagger(r') \, \psi(r') \, \psi(r) + \frac{1}{2} \int dr \, V(0) \, \psi^\dagger(r) \, \psi(r')$$

$$= \int dr \, \frac{\nabla \psi^\dagger(r) \, \nabla \psi(r)}{2 \, m} + \frac{1}{2} \int dr \, dr' \, V(\,|\,r - r'\,|\,) \, \psi^\dagger(r) \, \psi^\dagger(r') \, \psi(r') \, \psi(r) + \frac{V(0) \, N}{2}$$

(8.4)

We can drop the last term, which is a constant. (It is canceled by the potential energy from the nucleons). This procedure is known as "normal order", i.e. putting creation operators on the left and annihilation operators on the right.

$$H = \int dr \, \frac{\nabla \psi^\dagger(r) \, \nabla \psi(r)}{2 \, m} + \frac{1}{2} \int dr \, dr' \, V(\,|\,r - r'\,|\,) \, \psi^\dagger(r) \, \psi^\dagger(r') \, \psi(r') \, \psi(r)$$

(8.5)

8.1.5. Equation of motion for operators

In general, for an operator X

$$i \, \frac{\partial X(t)}{\partial t} = [X(t), \, H(t)]$$

(8.6)

For operator ψ, we have

$$i \, \frac{\partial \psi(r_0, t)}{\partial t} = [\psi(r_0), \, H] = \int dr \, \frac{\nabla \, \delta(r - r_0) \, \nabla \psi(r)}{2 \, m} +$$

$$\frac{1}{2} \int dr \, dr' \, \delta(r - r_0) \, V(\,|\,r - r'\,|\,) \, \psi^\dagger(r') \, \psi(r') \, \psi(r) + \frac{1}{2} \int dr \, dr' \, V(\,|\,r - r'\,|\,) \, \delta(r' - r_0) \, \psi^\dagger(r) \, \psi(r') \, \psi(r)$$

$$= -\frac{\nabla^2 \psi(r_0, t)}{2 \, m} + \int dr \, V(\,|\,r_0 - r\,|\,) \, \psi^\dagger(r) \, \psi(r) \, \psi(r_0)$$

(8.7)

Similarly, for the conjugate operator $\psi^\dagger$, we have

$$-i \, \frac{\partial \psi^\dagger(r_0, t)}{\partial t} = -\frac{\nabla^2 \psi^\dagger(r_0, t)}{2 \, m} + \psi^\dagger(r_0) \int dr \, V(\,|\,r_0 - r\,|\,) \, \psi^\dagger(r) \, \psi(r)$$

(8.8)

8.1.6. Equation of motion for correlation functions

Define

$$\tilde{G}(r_1, \, t_1; \, r_2 \, t_2) = \frac{1}{i} \left\langle \psi(r_1, \, t_1) \, \psi^\dagger(r_2, \, t_2) \right\rangle$$

(8.9)

For systems with translational symmetry in space and time, $\tilde{G}(r_1, \, t_1; \, r_2 \, t_2) = \tilde{G}(r_1 - r_2, \, t_1 - t_2) = \tilde{G}(r, \, t)$.

Q: What is the equations of motion for $G(r, \, t)$?

$$\partial_t \tilde{G}(r.t) = \partial_{t_1} \frac{1}{i} \left\langle \psi(r_1, \, t_1) \, \psi^\dagger(r_2, \, t_2) \right\rangle = \left\langle -i \, \partial_{t_1} \, \psi(r_1, \, t_1) \, \psi^\dagger(r_2, \, t_2) \right\rangle$$

$$= -\left\langle \left[-\frac{\nabla^2 \psi(r_1, \, t_1)}{2 \, m} + \int dr \, V(\,|\,r_0 - r\,|\,) \, \psi^\dagger(r) \, \psi(r) \, \psi(r_1, \, t_1) \right] \psi^\dagger(r_2, \, t_2) \right\rangle$$

(8.10)

$$= \frac{1}{2\,m}\,\partial_{r_1}{}^2 \left\langle \psi(r_1,\,t_1)\,\psi^\dagger(r_2,\,t_2) \right\rangle - V\left\langle \psi^\dagger\,\psi^\dagger\,\psi\,\psi \right\rangle = \frac{i}{2\,m}\,\nabla^2\,\tilde{G}(r,\,t) - V<\psi^\dagger\,\psi^\dagger\,\psi\,\psi>$$

If $V = 0$ (free systems), we have a closed partial differential equation for two fermion correlation functions.

$$i\,\partial_t\,\tilde{G}_0(r.t) + \frac{\nabla^2}{2\,m}\,\tilde{G}_0(r,\,t) = 0 \tag{8.11}$$

We can just solve this partial differential equation (with proper initial conditions and boundary conditions), and we get the correlation function.

However, for $V \neq 0$, the story is not as simple. We need to solve a more complicate equation.

$$i\,\partial_t\,\tilde{G}(r.t) + \frac{\nabla^2}{2\,m}\,\tilde{G}(r,\,t) = -i\,V<\psi^\dagger\,\psi^\dagger\,\psi\,\psi> \tag{8.12}$$

So we are having an Inhomogeneous partial differential equation.

$$i\,\partial_t\,\tilde{G}(r.t) + \frac{\nabla^2}{2\,m}\,\tilde{G}(r,\,t) = f(r,\,t) \tag{8.13}$$

Q: How do we solve Inhomogeneous partial differential equation?

A: Let's look at E&M

8.1.7. E&M: electric potential ϕ(r) for charge distribution ρ(r)

Gauss's law

$$\nabla \cdot \vec{E}(r) = \rho(r) \tag{8.14}$$

We know that

$$\vec{E}(r) = -\nabla\,\phi(r) \tag{8.15}$$

so

$$\nabla^2\,\phi(r) = \rho(r) \tag{8.16}$$

This is an Inhomogeneous equation. How do we solve it?

We first solve a different equation:

$$\nabla^2\,G(r,\,r_0) = \delta(r_0) \tag{8.17}$$

$G(r,\,r_0)$ is the electric potential for a point charge at position r_0. We know the solution of this equation, which is just the Coulomb's law

$$G(r,\,r_0) = \frac{1}{4\,\pi}\,\frac{1}{|\,r - r_0\,|} \tag{8.18}$$

Then, we know that

$$\phi(r) = \int d\,r_0\,G(r,\,r_0)\,\rho(r_0) = \int d\,r_0\,\frac{1}{4\,\pi}\,\frac{1}{|\,r - r_0\,|}\,\rho(r_0) \tag{8.19}$$

Mathematicians calls $G(r,\,r_0)$ Green's function. And this methods of solving inhomogeneous PDEs are known as the Green's function approach.

In general, one first change the inhomogeneous part with a delta function. Then, one solve this PDE first and the solution is known as the Green's function. Then one use this Integral to find the solution for the original inhomogeneous PDE.

8.1.8. A trick to get delta function: time ordering

Define time-ordered correlation functions (Green's functions)

$$G(r_1,\,t_1;\,r_2\,t_2) = \frac{1}{i}\,\left\langle \mathcal{T}\psi(r_1,\,t_1)\,\psi^\dagger(r_2,\,t_2) \right\rangle \tag{8.20}$$

Here $\mathcal{T}\psi^\dagger(r_1, t_1)\,\psi(r_2, t_2)$ is known as the time – ordered product.

$$\mathcal{T}\,\psi(r_1, t_1)\,\psi^\dagger(r_2, t_2) = \psi(r_1, t_1)\,\psi^\dagger(r_2, t_2) \quad \text{if } t_1 > t_2$$
$$= \pm\,\psi^\dagger(r_2, t_2)\,\psi(r_1, t_1) \quad \text{if } t_1 < t_2 \tag{8.21}$$

For bosons, we use the + sign and for fermions we use the - sign

In other words,

$$\mathcal{T}\,\psi(r_1, t_1)\,\psi^\dagger(r_2, t_2) = \psi(r_1, t_1)\,\psi^\dagger(r_2, t_2)\,\eta(t_1 - t_2) \pm \psi^\dagger(r_2, t_2)\,\psi(r_1, t_1)\,\eta(t_2 - t_1) \tag{8.22}$$

where $\eta(x)$ is the step function $\eta(\mathrm{x})=1$ for x>0 and $\eta(\mathrm{x})=0$ for x<0

Now, let's consider the EOM for $G(r, t)$

$$\partial_t \tilde{G}(r.t) = \partial_{t_1} \frac{1}{i} \left\langle \mathcal{T}\,\psi(r_1, t_1)\,\psi^\dagger(r_2, t_2) \right\rangle = -i\,\partial_{t_1} \left\langle \psi(r_1, t_1)\,\psi^\dagger(r_2, t_2)\,\eta(t_1 - t_2) \pm \psi^\dagger(r_2, t_2)\,\psi(r_1, t_1)\,\eta(t_2 - t_1) \right\rangle$$

$$= \left\langle -i\,\partial_{t_1}\psi(r_1, t_1)\,\psi^\dagger(r_2, t_2)\,\eta(t_1 - t_2) \right\rangle \pm \left\langle \psi^\dagger(r_2, t_2)\,(-i\,\partial_{t_1})\,\psi(r_1, t_1)\,\eta(t_2 - t_1) \right\rangle -$$
$$i\,\left\langle \psi(r_1, t_1)\,\psi^\dagger(r_2, t_2) \mp \psi^\dagger(r_2, t_2)\,\psi(r_1, t_1) \right\rangle \delta(t_1 - t_2)$$

$$= \frac{1}{2m}\,\partial_{r_1}{}^2 \left\langle \mathcal{T}\psi(r_1, t_1)\,\psi^\dagger(r_2, t_2) \right\rangle - V\left\langle \mathcal{T}\,\psi^\dagger\,\psi^\dagger\,\psi\,\psi \right\rangle - i\,\delta(r_1 - r_2)\,\delta(t_1 - t_2) \tag{8.23}$$

$$= \frac{i}{2m}\,\nabla^2 G(r, t) - V\left\langle \mathcal{T}\,\psi^\dagger\,\psi^\dagger\,\psi\,\psi \right\rangle - i\,\delta(r_1 - r_2)\,\delta(t_1 - t_2)$$

If $V = 0$

$$i\,\partial_t G_0(r.t) + \frac{1}{2m}\,\nabla^2 G_0(r, t) = \delta(r)\,\delta(t) \tag{8.24}$$

Bottom line, by time-ordering, we automatically got the Green's function.

$$\left(i\,\partial_t + \frac{\nabla^2}{2m}\right) G_0(r, t) = \delta(r)\,\delta(t) \tag{8.25}$$

$$\int dr\,dt\,\exp[-i\,k\,r + i\,\omega\,t]\left(i\,\partial_t + \frac{\nabla^2}{2m}\right) G_0(r, t) = \int dr\,dt\,\exp[-i\,k\,r + i\,\omega\,t]\,\delta(r)\,\delta(t) \tag{8.26}$$

$$\int dr\,dt\,\left(-i\,\partial_t + \frac{\nabla^2}{2m}\right)\exp[-i\,k\,r + i\,\omega\,t]\,G_0(r, t) = 1 \tag{8.27}$$

$$\left(\omega - \frac{k^2}{2m}\right)\int dr\,dt\,\exp[-i\,k\,r + i\,\omega\,t]\,G_0(r, t) = 1 \tag{8.28}$$

$$\left(\omega - \frac{k^2}{2m}\right) G_0(k, \omega) = 1 \tag{8.29}$$

$$G_0(k, \omega) = \frac{1}{\omega - \frac{k^2}{2m}} \tag{8.30}$$

More generally, for free particles (without interactions), $G_0 = \frac{1}{\omega - \epsilon_k}$

8.1.9. Note: there are many other reason to use $\mathcal{T}$.

(1) Path integral leads to $\mathcal{T}$ naturally.

(2) The evaluational operator $\mathcal{T}\exp[i\int dt\;H(t)\,t]$

(3) With $\mathcal{T}$, bosons and fermions are unified together. Same theory with two different boundary conditions.

8.2. Boundary condition and connections between different Green's functions

8.2.1. Other correlation functions and the boundary condition

Define

$$G^>(1, 2) = \frac{1}{i} \langle \psi(1) \psi^\dagger(2) \rangle \tag{8.31}$$

$$G^<(1, 2) = \pm \frac{1}{i} \langle \psi^\dagger(2) \psi(1) \rangle \tag{8.32}$$

For statistical average, we have Boltzmann factor $e^{\beta H}$. For time-evolution, we have the evolution operator $e^{i H t}$. It seems that inverse temperature β is just the imaginary time. Let's try this idea by allowing time to be complex.

For $G^>$

$$G^>(1, 2) = \frac{1}{i} \langle \psi(1) \psi^\dagger(2) \rangle = \frac{\text{Tr}\left\{\exp\left[-\beta(\hat{H} - \mu \hat{N})\right] \psi(r_1, t_1) \psi^\dagger(r_2, t_2)\right\}}{i \,\text{Tr}\left\{\exp\left[-\beta(\hat{H} - \mu \hat{N})\right]\right\}} = \tag{8.33}$$

$$\left(\text{Tr}\left\{\exp\left[-\beta(\hat{H} - \mu \hat{N})\right] \exp(i \hat{H} t_1) \psi(r_1) \exp(-i \hat{H} t_1) \exp(i \hat{H} t_2) \psi^\dagger(r_2) \exp(-i \hat{H} t_2)\right\}\right) \Big/ \left(i \,\text{Tr}\left\{\exp\left[-\beta(\hat{H} - \mu \hat{N})\right]\right\}\right)$$

If we use eigenenergy states to compute the sum

$$G^>(1, 2) =$$
$$\left(\sum_n \langle n | \left\{\exp\left[-\beta(\hat{H} - \mu \hat{N})\right] \exp(i \hat{H} t_1) \psi(r_1) \exp(-i \hat{H} t_1) \exp(i \hat{H} t_2) \psi^\dagger(r_2) \exp(-i \hat{H} t_2)\right\} | n \rangle\right) \Big/ \left(i \,\text{Tr}\left\{\exp\left[-\beta(\hat{H} - \mu \hat{N})\right]\right\}\right) =$$
$$\exp[\beta \mu N] \left(\sum_n \exp[E_n(i t_1 - \beta)] \langle n | \psi(r_1) \exp(-i \hat{H} t_1) \exp(i \hat{H} t_2) \psi^\dagger(r_2) | n \rangle \exp(-i E_n t_2)\right) \Big/ \left(i \,\text{Tr}\left\{\exp\left[-\beta(\hat{H} - \mu \hat{N})\right]\right\}\right) = \tag{8.34}$$
$$\exp[\beta \mu N] \frac{\sum_n \exp[E_n(i t_1 - i t_2 - \beta)] \langle n | \psi(r_1) \exp(-i \hat{H} t_1) \exp(i \hat{H} t_2) \psi^\dagger(r_2) | n \rangle}{i \,\text{Tr}\left\{\exp\left[-\beta(\hat{H} - \mu \hat{N})\right]\right\}}$$

Because E_n has no upper bound, as $E_n \to +\infty$, to keep the factor $\exp[E_n(i t_1 - i t_2 - \beta)]$ converge, we need to have `Re (i t₁ - i t₂ - β) < 0`, In other words, $\text{Im}(t_1 - t_2) > -\beta$.

In the same time, using $1 = \sum_m | m \rangle \langle m |$, we have

$$G^>(1, 2) =$$
$$\exp[\beta \mu N] \left(\sum_n \exp[E_n(i t_1 - i t_2 - \beta)] \langle n | \psi(r_1) \sum_m | m \rangle \langle m | \exp(-i \hat{H} t_1) \exp(i \hat{H} t_2) \psi^\dagger(r_2) | n \rangle\right) \Big/ \left(i \,\text{Tr}\left\{\exp\left[-\beta(\hat{H} - \mu \hat{N})\right]\right\}\right) = \tag{8.35}$$
$$\exp[\beta \mu N] \left(\sum_{n,m} \exp[E_n(i t_1 - i t_2 - \beta)] \langle n | \psi(r_1) | m \rangle \langle m | \psi^\dagger(r_2) | n \rangle \exp[i E_m(t_2 - t_1)]\right) \Big/ \left(i \,\text{Tr}\left\{\exp\left[-\beta(\hat{H} - \mu \hat{N})\right]\right\}\right)$$

To make sure that $E_m \to +\infty$ shows no singularity, $\text{Im}(t_1 - t_2) < 0$.

So, $0 > \text{Im}(t_1 - t_2) > -\beta$

For $G^<$, we have $0 < \text{Im}(t_1 - t_2) < \beta$

Let's come back to $G^>$

$$G^>(r_1, t_1 - i \beta; r_2, t_2) =$$
$$\frac{\text{Tr}\left\{\exp\left[-\beta(\hat{H} - \mu \hat{N})\right] \exp(\beta \hat{H}) \exp(i \hat{H} t_1) \psi(r_1) \exp(-i \hat{H} t_1) \exp(-\beta \hat{H}) \exp(i \hat{H} t_2) \psi^\dagger(r_2) \exp(-i \hat{H} t_2)\right\}}{i \,\text{Tr}\left\{\exp\left[-\beta(\hat{H} - \mu \hat{N})\right]\right\}} =$$
$$\frac{\text{Tr}\left\{\exp[\beta \mu \hat{N}] \exp(i \hat{H} t_1) \psi(r_1) \exp(-i \hat{H} t_1) \exp(-\beta \hat{H}) \exp(i \hat{H} t_2) \psi^\dagger(r_2) \exp(-i \hat{H} t_2)\right\}}{i \,\text{Tr}\left\{\exp\left[-\beta(\hat{H} - \mu \hat{N})\right]\right\}} =$$
$$\frac{\text{Tr}\left\{\exp(i \hat{H} t_1) \psi(r_1) \exp\left[\beta \mu (\hat{N} - 1)\right] \exp(-i \hat{H} t_1) \exp(-\beta \hat{H}) \exp(i \hat{H} t_2) \psi^\dagger(r_2) \exp(-i \hat{H} t_2)\right\}}{i \,\text{Tr}\left\{\exp\left[-\beta(\hat{H} - \mu \hat{N})\right]\right\}} = \tag{8.36}$$

$$e^{-\beta\mu} \frac{\mathrm{Tr}\left\{\exp\left(i\,\hat{H}\,t_1\right)\psi(r_1)\exp\left(-i\,\hat{H}\,t_1\right)\exp\left[-\beta\left(\hat{H}-\mu\,\hat{N}\right)\right]\exp\left(i\,\hat{H}\,t_2\right)\psi^\dagger(r_2)\exp\left(-i\,\hat{H}\,t_2\right)\right\}}{i\,\mathrm{Tr}\left\{\exp\left[-\beta\left(\hat{H}-\mu\,\hat{N}\right)\right]\right\}} =$$

$$e^{-\beta\mu} \frac{\mathrm{Tr}\left\{\psi(1)\exp\left[-\beta\left(\hat{H}-\mu\,\hat{N}\right)\right]\psi^\dagger(2)\right\}}{i\,\mathrm{Tr}\left\{\exp\left[-\beta\left(\hat{H}-\mu\,\hat{N}\right)\right]\right\}} = e^{-\beta\mu} \frac{\mathrm{Tr}\left\{\exp\left[-\beta\left(\hat{H}-\mu\,\hat{N}\right)\right]\psi^\dagger(2)\,\psi(1)\right\}}{i\,\mathrm{Tr}\left\{\exp\left[-\beta\left(\hat{H}-\mu\,\hat{N}\right)\right]\right\}} = \pm e^{-\beta\mu}\,G^<(1,\,2)$$

So, $e^{\beta\mu}\,G^>(r_1,\,t_1-i\,\beta;\,r_2,\,t_2) = \pm\,G^<(r_1,\,t_1;\,r_2,\,t_2)$ (8.37)

8.2.2. BC for the time-ordered Green's function.

The time-ordered Green's function can be defined along the imaginary axis also

$$\mathcal{T}\,\psi(r_1,\,t_1)\,\psi^\dagger(r_2,\,t_2) = \psi(r_1,\,t_1)\,\psi^\dagger(r_2,\,t_2) \quad \text{if } \mathrm{Im}(t_1) > \mathrm{Im}(t_2)$$
$$= \pm\psi^\dagger(r_2,\,t_2)\,\psi(r_1,\,t_1) \quad \text{if } \mathrm{Im}(t_1) < \mathrm{Im}(t_2)$$
(8.38)

So,

$$\mathcal{T}\,G = G^> \quad \text{if } \mathrm{Im}(t_1) < \mathrm{Im}(t_2)$$
$$\pm\,G^< \quad \text{if } \mathrm{Im}(t_1) > \mathrm{Im}(t_2)$$
(8.39)

It is easy to check that for $G(r,\,t)$, we have

$$G(r,\,t) = \pm\,e^{\beta\mu}\,G(r,\,t-i\,\beta)$$ (8.40)

This is a boundary condition for imaginary time. At $\mu=0$, Bosons have a periodic BC, and fermions have an anti-periodic BC.

8.2.3. the k,ω space

Define

$$G^>(k,\,\omega) = i \int dr\,dt\,e^{-i\,k\,r+i\,\omega\,t}\,G^>(r,\,t) = \left\langle \psi_{k,\omega}\,\psi_{k,\omega}^\dagger \right\rangle$$
$$G^<(k,\,\omega) = \pm i \int dr\,dt\,e^{-i\,k\,r+i\,\omega\,t}\,G^<(r,\,t) = \left\langle \psi_{k,\omega}^\dagger\,\psi_{k,\omega} \right\rangle$$
(8.41)

$$G^<(k,\,\omega) = \pm i \int dr\,dt\,e^{-i\,k\,r+i\,\omega\,t}\,G^<(r,\,t) =$$
$$i\,e^{\beta\mu} \int dr\,dt\,e^{-i\,k\,r+i\,\omega\,t}\,G^>(r,\,t-i\,\beta) = i\,e^{\beta\mu} \int dr\,dt'\,e^{-i\,k\,r+i\,\omega\,(t'+i\,\beta)}\,G^>(r,\,t') = e^{-\beta(\omega-\mu)}\,G^>(k,\,\omega)$$
(8.42)

Define the "spectral function"

$$A(k,\,\omega) = G^>(k,\,\omega) \mp G^<(k,\,\omega)$$ (8.43)

Because $G^<(k,\,\omega) = e^{-\beta(\omega-\mu)}\,G^>(k,\,\omega)$

$$A(k,\,\omega) = G^>(k,\,\omega) \mp G^<(k,\,\omega) = G^>(k,\,\omega) \mp e^{-\beta(\omega-\mu)}\,G^>(k,\,\omega)$$ (8.44)

So

$$G^>(k,\,\omega) = \frac{A(k,\,\omega)}{1 \mp e^{-\beta(\omega-\mu)}} = A(k,\,\omega)\,\frac{e^{\beta(\omega-\mu)}}{e^{\beta(\omega-\mu)} \mp 1} = A(k,\,\omega)\left[1 \pm \frac{1}{e^{\beta(\omega-\mu)} \mp 1}\right] = A(k,\,\omega)[1 \pm f(\omega)]$$ (8.45)

$$G^<(k,\,\omega) = e^{-\beta(\omega-\mu)}\,G^>(k,\,\omega) = e^{-\beta(\omega-\mu)}\,\frac{A(k,\,\omega)}{1 \mp e^{-\beta(\omega-\mu)}} = A(k,\,\omega)\,\frac{1}{e^{\beta(\omega-\mu)} \mp 1} = A(k,\,\omega)\,f(\omega)$$ (8.46)

Here $f(\omega)$ is the boson/fermion distribution function.

8.2.4. G in the k,ω space part I: the Matsubara frequencies

Remember that $G(r,\,t) = \pm\,e^{\beta\mu}\,G(r,\,t-i\,\beta)$, which is (almost) a PBC (anti-PBC) for t. We know that boundary conditions implies quantization (discrete ω).

For example, for a periodic function $f(t) = f(t + T)$, we know that in the frequency space, $f(\omega)$ is defined only on a discrete set of frequency points, $\omega = 2\pi n/T$. For anti-PBC, $f(t) = -f(t + T)$, $\omega = (2n + 1)\pi/T$.

For $G(r, t) = \pm e^{\beta\mu} G(r, t - i\beta)$, if we go to the k,$\omega$ space and define

$$G(r, t) = \int dk \, d\omega \, e^{ikr - i\omega t} \, G(k, \omega) \tag{8.47}$$

The condition $G(r, t) = \pm e^{\beta\mu} G(r, t - i\beta)$ implies that

$$\int dk \, d\omega \, e^{ikr - i\omega t} \, G(k, \omega) = \pm e^{\beta\mu} \, e^{-\omega\beta} \int dk \, d\omega \, e^{ikr - i\omega t} \, G(k, \omega) \tag{8.48}$$

So, $\pm e^{\beta\mu} \, e^{-\omega\beta} = 1$

$$e^{\beta(\omega - \mu)} = \pm 1 \tag{8.49}$$

For bosons,

$$\beta(\omega - \mu) = 2n\pi i \tag{8.50}$$

$$\omega = \frac{2n\pi i}{\beta} + \mu \tag{8.51}$$

Following the historical convention, we write $\omega = i\omega_n$

$$\omega_n = \frac{2n\pi}{\beta} - i\mu \tag{8.52}$$

These discrete frequency points ω_n are known as the Matsubara frequencies

For fermions,

$$\beta(\omega - \mu) = (2n + 1)\pi i \tag{8.53}$$

$$\omega = \frac{(2n + 1)\pi i}{\beta} + \mu \tag{8.54}$$

$$\omega_n = \frac{2n + 1}{\beta}\pi - i\mu \tag{8.55}$$

8.2.5. G in the k,ω space part II: analytic continuation

After we find $G(k, i\omega_n)$, where ω_n takes discrete values, now let's define a new function $G(k, z)$ by simply replacing the discrete number $i\omega_n$ into a complex number z that various continuously. For the function $G(k, z)$, it is a function well-defined at every point on the complex z plane (there may be some singularity points). At the Matsubara frequencies, this new function $G(k, z)$ coincides with $G(k, i\omega_n)$. This procedure is known as analytic continuation. This new function is very useful. Here, I will show you that by defining $G(k, z)$, we can find a very easy way to get $G^>$ and $G^<$ from G. Later, we will use this $G(k, z)$ to compute $G(k, i\omega_n)$.

$$G(k, i\omega_n) = \int_0^{-i\beta} dt \, e^{-\omega_n t} \, G(k, t) =$$

$$\int_0^{-i\beta} dt \, e^{-\omega_n t} \, G^>(k, t) = \int_0^{-i\beta} dt \, e^{-\omega_n t} \int \frac{d\omega}{2\pi i} e^{-i\omega t} \, G^>(k, \omega) = \int \frac{d\omega}{2\pi} \frac{e^{(-i\omega - \omega_n)(-i\beta)} - 1}{\omega - i\omega_n} \frac{A(k, \omega)}{1 \mp e^{-\beta(\omega - \mu)}} = \tag{8.56}$$

$$\int \frac{d\omega}{2\pi} \frac{1 - e^{-(\omega - i\omega_n)\beta}}{\omega - i\omega_n} \frac{A(k, \omega)}{1 \mp e^{-\beta(\omega - \mu)}} = \int \frac{d\omega}{2\pi} \frac{1 \mp e^{-(\omega - \mu)\beta}}{i\omega_n - \omega} \frac{A(k, \omega)}{1 \mp e^{-\beta(\omega - \mu)}} = \int \frac{d\omega}{2\pi} \frac{A(k, \omega)}{i\omega_n - \omega}$$

Therefore,

$$G(k, z) = \int \frac{d\omega}{2\pi} \frac{A(k, \omega)}{z - \omega} \tag{8.57}$$

Now, we substitute z by $\omega + i\delta$

$$G(k, \omega + i\epsilon) = \int \frac{d\Omega}{2\pi} \frac{A(k, \Omega)}{\omega + i\epsilon - \Omega} \tag{8.58}$$

$$\text{Im}[G(k, \omega + i\epsilon)] = \frac{G(k, \omega + i\epsilon) - G(k, \omega - i\epsilon)}{2i} = \frac{1}{2i} \int \frac{d\Omega}{2\pi} \left[\frac{A(k, \Omega)}{\omega + i\epsilon - \Omega} - \frac{A(k, \Omega)}{\omega - i\epsilon - \Omega} \right] =$$

$$\frac{1}{2i} \int \frac{d\Omega}{2\pi} \frac{A(k, \Omega)}{(\omega - \Omega)^2 + \epsilon^2} (-2i\epsilon) = -\int \frac{d\Omega}{2\pi} \frac{\epsilon}{(\omega - \Omega)^2 + \epsilon^2} A(k, \Omega) = -\frac{1}{2} \int d\Omega \, \delta(\omega - \Omega) A(k, \Omega) = -\frac{1}{2} A(k, \omega) \tag{8.59}$$

So,

$$A(k, \omega) = -2 \, \text{Im}[G(k, \omega + i\epsilon)] \tag{8.60}$$

8.2.6. Example: free particles

$$i \, \partial_t G_0(r.t) + \frac{1}{2m} \nabla^2 G_0(r, t) = \delta(r) \, \delta(t) \tag{8.61}$$

in the k,ω space

$$\left(i\omega_n - \frac{k^2}{2m} \right) G_0(k, i\omega_n) = 1 \tag{8.62}$$

$$G_0(k, i\omega_n) = \frac{1}{i\omega_n - \frac{k^2}{2m}} \tag{8.63}$$

More generic case: dispersion relation $\epsilon(k)$

$$G_0(k, i\omega_n) = \frac{1}{i\omega_n - \epsilon(k)} \tag{8.64}$$

Analytic continuation:

$$G_0(k, z) = \frac{1}{z - \epsilon(k)} \tag{8.65}$$

the spectral function:

$$A(k, \omega) = -2 \, \text{Im}[G(k, \omega + i\epsilon)] = -2 \, \text{Im}\left[\frac{1}{\omega + i\epsilon - \epsilon(k)} \right] = \frac{2\epsilon}{(\omega - \epsilon_k)^2 + \epsilon^2} = 2\pi \, \delta(\omega - \epsilon_k) \tag{8.66}$$

particular number:

$$\langle n_k(t) \rangle = \left\langle \psi_k^\dagger(t) \, \psi_k(t) \right\rangle = G^<(k, t - t) = \int \frac{d\omega}{2\pi} G^<(k, \omega) = \int \frac{d\omega}{2\pi} 2\pi \, \delta(\omega - \epsilon_k) f(\omega) = f(\epsilon_k) \tag{8.67}$$

$$\langle 1 \pm n_k(t) \rangle = \left\langle \psi_k(t) \, \psi_k^\dagger(t) \right\rangle = G^>(k, t - t) = \int \frac{d\omega}{2\pi} G^>(k, \omega) = \int \frac{d\omega}{2\pi} 2\pi \, \delta(\omega - \epsilon_k)[1 \pm f(\omega)] = 1 \pm f(\epsilon_k) \tag{8.68}$$

8.2.7. Interacting particles.

Once we found $G(k, i\omega_n)$,

We can get $A(k, \omega)$ using

$$A(k, \omega) = -2 \, \text{Im}[G(k, \omega + i\epsilon)] \tag{8.69}$$

Then we can get all other correlation functions like $G^<(k, t - t)$ and $G^>(k, t - t)$ using

$$G^>(k, \omega) = \frac{A(k, \omega)}{1 \mp e^{-\beta(\omega - \mu)}} = A(k, \omega) \frac{e^{\beta(\omega - \mu)}}{e^{\beta(\omega - \mu)} \mp 1} = A(k, \omega) \left[1 \pm \frac{1}{e^{\beta(\omega - \mu)} \mp 1} \right] = A(k, \omega)[1 \pm f(\omega)] \tag{8.70}$$

$$G^<(k,\,\omega) = e^{-\beta(\omega-\mu)} \, G^>(k,\,\omega) = e^{-\beta(\omega-\mu)} \, \frac{A(k,\,\omega)}{1 \mp e^{-\beta(\omega-\mu)}} = A(k,\,\omega) \, \frac{1}{e^{\beta(\omega-\mu)} \mp 1} = A(k,\,\omega) \, f(\omega) \tag{8.71}$$

8.2.8. Other particles (free particles only in this section)

For free particles, the Green's function has the same equation of motion as the wavefunction, but with an extra δ-function.

For particles that follow the Schrödinger equation,

$$i \, \partial_t \psi = -\frac{\nabla^2}{2\,m} \psi \tag{8.72}$$

$$\left(i \, \partial_t + \frac{\nabla^2}{2\,m}\right) \psi = 0 \tag{8.73}$$

The Green's function has the same EMO, but with a δ-function on the r.h.s.

$$\left(i \, \partial_t + \frac{\nabla^2}{2\,m}\right) G(r,\,t) = \delta(r) \, \delta(t) \tag{8.74}$$

If we go to k-space,

$$(\omega - \epsilon_k) \, G(k,\,\omega) = 1 \tag{8.75}$$

$$G(k,\,\omega) = \frac{1}{\omega - \epsilon_k} \tag{8.76}$$

In some sense, the Green's function is just the inverse of the EOM

$$G(r,\,t) = \left(i \, \partial_t + \frac{\nabla^2}{2\,m}\right)^{-1} \tag{8.77}$$

8.3. Feynman diagram

8.3.1. multi-particle Green's functions

Single particle Green's function:

$$G(1, 1') = \left(\frac{1}{i}\right) \langle \mathcal{T} \psi(1) \, \psi^\dagger(1') \rangle \tag{8.72}$$

Two-particle Green's function:

$$G_2(1, 2; 1', 2') = \left(\frac{1}{i}\right)^2 \langle \mathcal{T} \psi(1) \, \psi(2) \, \psi^\dagger(2') \, \psi^\dagger(1') \rangle \tag{8.73}$$

three-particle Green's function:

$$G_3(1, 2, 3; 1', 2', 3') = \left(\frac{1}{i}\right)^3 \langle \mathcal{T} \psi(1) \, \psi(2) \, \psi(3) \, \psi^\dagger(3') \, \psi^\dagger(2') \, \psi^\dagger(1') \rangle \tag{8.74}$$

n-particle Green's function:

$$G_n(1, 2, \ldots, n; 1', 2', \ldots, n') = \left(\frac{1}{i}\right)^n \langle \mathcal{T} \psi(1) \, \psi(2) \ldots \psi(n) \, \psi^\dagger(n') \ldots \psi^\dagger(2') \, \psi^\dagger(1') \rangle \tag{8.75}$$

8.3.2. the equations of motion of the Green's functions

$$i \frac{\partial \psi(r_0, t)}{\partial t} = [\psi(r_0), H] = \int dr \, \frac{\nabla \delta(r - r_0) \nabla \psi(r)}{2 m} +$$

$$\frac{1}{2} \int dr \, dr' \, \delta(r - r_0) \, V(\,|\,r - r'\,|) \, \psi^\dagger(r') \, \psi(r') \, \psi(r) + \frac{1}{2} \int dr \, dr' \, V(\,|\,r - r'\,|) \, \delta(r' - r_0) \, \psi^\dagger(r) \, \psi(r') \, \psi(r) \tag{8.76}$$

$$= -\frac{\nabla^2 \psi(r_0, t)}{2 m} + \int dr \, V(\,|\,r_0 - r\,|) \, \psi^\dagger(r) \, \psi(r) \, \psi(r_0)$$

$$\left(i \, \partial_t + \frac{1}{2 m} \nabla^2\right) G(1, 2) = \delta(r_1 - r_2) \, \delta(t_1 - r_2) + \frac{1}{i} \int dr \, V(\,|\,r_0 - r\,|) \, \langle \mathcal{T} \psi^\dagger(r_1, t_1) \, \psi(r_1, t_1) \, \psi(r_0, t_1) \, \psi^\dagger(r_2, t_2) \rangle \tag{8.77}$$

$$\left(i \, \partial_t + \frac{1}{2 m} \nabla^2\right) G(1, 1') = \delta(1 - 1') + i \int dr \, V(\,|\,r_1 - r_2\,|) \left(\frac{1}{i}\right)^2 \langle \mathcal{T} \left[\psi^\dagger(r_2, t_1) \, \psi(r_2, t_1) \, \psi(r_1, t_1)\right] \psi^\dagger(r_{1'}, t_{1'}) \rangle =$$

$$\delta(1 - 1') + i \int dr \, V(\,|\,r_1 - r_2\,|) \left(\frac{1}{i}\right)^2 \langle \mathcal{T} \psi^\dagger(r_2, t_1 + \delta) \, \psi(r_2, t_1) \, \psi(r_1, t_1) \, \psi^\dagger(r_{1'}, t_{1'}) \rangle =$$

$$\delta(1 - 1') \pm i \int dr \, V(\,|\,r_1 - r_2\,|) \left(\frac{1}{i}\right)^2 \langle \mathcal{T} \psi(r_1, t_1) \, \psi(r_2, t_1) \, \psi^\dagger(r_2, t_1 + \delta) \, \psi^\dagger(r_{1'}, t_{1'}) \rangle = \tag{8.78}$$

$$\delta(1 - 1') \pm i \int dr \, V(\,|\,r_1 - r_2\,|) \, G_2(1, 2; 1', 2^+) \,|_{t_1 = t_2}$$

Here 2^+ means that the time argument for 2^+ is slightly larger than 2, to keep the operators in the right order ($t_2^+ = t_2 + \delta$).

$$\left(i \, \partial_t + \frac{1}{2 m} \nabla^2\right) G(1, 1') = \delta(1 - 1') \pm i \int dr_2 \, V(\,|\,r_1 - r_2\,|) \, G_2(1, 2; 1', 2^+) \,|_{t_1 = t_2} \tag{8.79}$$

This equation tells us that in order to get $G(1, 2)$, we need to know $G_2(1, 2; 1', 2')$. So we need to write down the EOM for G_2

$$\left(i \, \partial_t + \frac{1}{2 m} \nabla^2\right) G_2(1, 2; 1', 2') = \delta(1 - 1') \, G(2, 2') \pm \delta(1 - 2') \, G(2, 1') \pm i \int dr_2 \, V(\,|\,r_1 - r_2\,|) \, G_3(1, 2, 3; 1', 2', 3^+) \,|_{t_1 = t_3} \tag{8.80}$$

This equation tells us that in order to get G_2, we need to know G_3.

Repeat the same procedure, we find that if we want to know G_n, we need to know G_{n+1}.

$$\left(i\,\partial_t + \frac{1}{2\,m}\,\nabla^2\right) G_n(1,\,2,\,\ldots,\,n;\,1',\,2',\,\ldots,\,n') = \delta(1-1')\,G_{n-1}(2,\,\ldots,\,n;\,2',\,\ldots,\,n') \pm \delta(1-2')\,G_{n-1}(2,\,\ldots,\,n;\,1',\,3',\,\ldots,\,n') +$$

$$\ldots \pm i \int d\,r_2 \; V(\mid r_1 - r_2 \mid)\, G_{n+1}(1,\,2,\,\ldots,\,n+1;\,1',\,2',\,\ldots,\,n+1)\mid_{t_1 = t_{n+1}} \tag{8.81}$$

So we cannot get a close set of equations. In order words, there is no way to solve these equations.

8.3.3. How to solve this equation? A: The perturbation theory

$G(1,\,1')$ depends on interaction strength V. Let's expand G as a power series of V

$$G(1,\,1') = G^{(0)}(1,\,1) + O(V) + O\big(V^2\big) + O\big(V^3\big) + \ldots \tag{8.82}$$

Notice that G_{n-1} is related to $V \times G_n$. Therefore, to get G up to the order of $O(V^n)$, we just need to keep G_2 to the order of $O(V^{n-1})$, and G_3 to $O(V^{n-2})\ldots G_{n+1}$ to $O(V^0)$ and set $G_{n+2} = 0$

8.3.4. the zeroth-order approximation (free-particle approximation, or say non-interacting approximation)

If we want to get G to the zeroth order, we need to set $G_2 = 0$

$$\left(i\,\partial_{t_1} + \frac{1}{2\,m}\,\nabla_{r_1}{}^2\right) G^{(0)}(1,\,1') = \delta(1-1') \tag{8.83}$$

One equation, one unknown. It can be solved easily.

$$G^{(0)}(k,\,i\omega_n) = \frac{1}{i\omega_n - \frac{k^2}{2\,m}} \tag{8.84}$$

8.3.5. the first-order approximation (the Hartree-Fock approximation)

If we want to get G to the first order, we need to set $G_3 = 0$ and keep G_2 to $O(V^0)$

$$\left(i\,\partial_{t_1} + \frac{1}{2\,m}\,\nabla_{r_1}{}^2\right) G(1,\,1') = \delta(1-1') \pm i \int d\,r_2 \; V(\mid r_1 - r_2 \mid)\, G_2(1,\,2;\,1',\,2^+)\mid_{t_1 = t_2} \tag{8.85}$$

$$\left(i\,\partial_{t_1} + \frac{1}{2\,m}\,\nabla_{r_1}{}^2\right) G_2(1,\,2;\,1',\,2') = \delta(1-1')\,G(2,\,2') \pm \delta(1-2')\,G(2,\,1') + O(V\,G_3) = \delta(1-1')\,G^{(0)}(2,\,2') \pm \delta(1-2')\,G^{(0)}(2,\,1') \tag{8.86}$$

two equations and two unknowns. The solution for the second equation is very simple:

$$G_2(1,\,2;\,1',\,2') = G^{(0)}(1,\,1')\,G^{(0)}(2,\,2') \pm G^{(0)}(1,\,2')\,G^{(0)}(2,\,1') \tag{8.87}$$

Let's check it

$$\left(i\,\partial_{t_1} + \frac{1}{2\,m}\,\nabla_{r_1}{}^2\right) G^{(0)}(1,\,1')\,G^{(0)}\big((2,\,2') = \delta(1-1')\,G^{(0)}((2,\,2') \tag{8.88}$$

$$\left(i\,\partial_{t_1} + \frac{1}{2\,m}\,\nabla_{r_1}{}^2\right) G^{(0)}(1,\,2')\,G^{(0)}\big((2,\,1') = \delta(1,\,2')\,G^{(0)}((2,\,1') \tag{8.89}$$

So we can go back to the first equation to get G.

$$\left(i\,\partial_{t_1} + \frac{1}{2\,m}\,\nabla_{r_1}{}^2\right) G(1,\,1') =$$

$$\delta(1-1') \pm i \int d\,r_2 \; V(\mid r_1 - r_2 \mid)\, G^{(0)}(1,\,1')\,G^{(0)}(2,\,2^+)\mid_{t_1 = t_2} + i \int d\,r_2 \; V(\mid r_1 - r_2 \mid)\, G^{(0)}(1,\,2^+)\,G^{(0)}(2,\,1')\mid_{t_1 = t_2} \tag{8.90}$$

$$\left(i\,\partial_{t_1} + \frac{1}{2\,m}\,\nabla_{r_1}{}^2\right) G(1,\,1') = \delta(1-1') + f(1,\,1') \tag{8.91}$$

This equation can be separated into two equations:

$$\left(i\,\partial_{t_1}+\frac{1}{2\,m}\,\nabla_{r_1}{}^2\right)G^{(0)}(1,\,1') = \delta(1-1')\tag{8.92}$$

$$\left(i\,\partial_{t_1}+\frac{1}{2\,m}\,\nabla_{r_1}{}^2\right)G^{(1)}(1,\,1') = f(1,\,1')\tag{8.93}$$

and $G = G^{(0)} + G^{(1)}$.

The first equation have been solved before. It is just the zeroth-order equation (the free theory)

$$\left(i\,\partial_{t_1}+\frac{1}{2\,m}\,\nabla_{r_1}{}^2\right)G^{(0)}(1,\,1') = \delta(1-1')\tag{8.94}$$

For the second equation, the solution is straightforward using the Green's function technique

$$G^{(1)}(1,\,1') = \int d\,r_3 \int d\,t_3\,G^{(0)}(1,\,3)\,f(3,\,1')\tag{8.95}$$

Let's check this

$$\left(i\,\partial_{t_1}+\frac{1}{2\,m}\,\nabla_{r_1}{}^2\right)\int d\,r_3 \int d\,t_3\,G^{(0)}(1,\,3)\,f(3,\,1') = \int d\,r_3 \int d\,t_3\,\delta(1-3)\,f(3,\,1') = f(1,\,1')\tag{8.96}$$

So

$$G^{(1)}(1,\,1') = \int d\,r_3 \int d\,t_3\,G^{(0)}(1,\,3)\,f(3,\,1') = \pm i\int d\,r_3 \int d\,t_3 \int d\,r_2\,G^{(0)}(1,\,3)\,V(\mid r_3 - r_2 \mid)\,G^{(0)}(3,\,1')\,G^{(0)}(2,\,2^+)\mid_{t_3=t_2}$$
$$+i\int d\,r_3 \int d\,t_3 \int d\,r_2\,G^{(0)}(1,\,3)\,V(\mid r_3 - r_2 \mid)\,G^{(0)}(3,\,2^+)\,G^{(0)}(2,\,1')\mid_{t_3=t_2}\tag{8.97}$$

For $G(1,\,1')$

$$G(1,\,1') = G^{(0)}(1,\,1') \pm i\int d\,r_3 \int d\,t_3 \int d\,r_2\,G^{(0)}(1,\,3)\,V(\mid r_3 - r_2 \mid)\,\delta(t_3-t_2)\,G^{(0)}(3,\,1')\,G^{(0)}(2,\,2^+) +$$
$$i\int d\,r_3 \int d\,t_3 \int d\,r_2\,G^{(0)}(1,\,3)\,V(\mid r_3 - r_2 \mid)\,\delta(t_3-t_2)\,G^{(0)}(3,\,2^+)\,G^{(0)}(2,\,1')\mid_{t_3=t_2}\tag{8.98}$$

The first term is the free propagator, the second therm is known as the Hartree term and the last term is the Fock term.

8.3.6. second order:

If we want to get G to the order of $O(V^2)$, we set $G_4 = 0$, keep G_3 to $O(V^0)$ and G_2 to $O(V^1)$

$$\left(i\,\partial_{t_1}+\frac{1}{2\,m}\,\nabla_{r_1}{}^2\right)G(1,\,1') = \delta(1-1') \pm i\int d\,r_2\,V(\mid r_1 - r_2 \mid)\,G_2(1,\,2;\,1',\,2^+)\mid_{t_1=t_2}\tag{8.99}$$

$$\left(i\,\partial_t+\frac{1}{2\,m}\,\nabla^2\right)G_2(1,\,2;\,1',\,2') = \delta(1-1')\,G(2,\,2') + \delta(1-2')\,G(2,\,1') \pm i\int d\,r_2\,V(\mid r_1 - r_2 \mid)\,G_3(1,\,2,\,3;\,1',\,2',\,3^+)\mid_{t_1=t_3}\tag{8.100}$$

$$\left(i\,\partial_{t_1}+\frac{1}{2\,m}\,\nabla_{r_1}{}^2\right)G_3(1,\,2,\,3;\,1',\,2',\,3') = \delta(1-1')\,G_2(2,\,3;\,2',\,3') \pm \delta(1-2')\,G_2(2,\,3;\,1',\,3') + \delta(1-3')\,G_2(2,\,3;\,1',\,2')\tag{8.101}$$

The last equation give us (keep only $O(V^0)$ terms)

$$G_3(1,\,2,\,3;\,1',\,2',\,3') = G^{(0)}(1,\,1')\,G^{(0)}(2,\,2')\,G^{(0)}(3,\,3') \pm G^{(0)}(1,\,1')\,G^{(0)}(2,\,3')\,G^{(0)}(2,\,3') \pm G^{(0)}(1,\,2')\,G^{(0)}(2,\,1')\,G^{(0)}(3,\,3') +$$
$$G^{(0)}(1,\,2')\,G^{(0)}(2,\,3')\,G^{(0)}(3,\,1') + G^{(0)}(1,\,3')\,G^{(0)}(2,\,1')\,G^{(0)}(3,\,2') \pm G^{(0)}(1,\,3')\,G^{(0)}(2,\,2')\,G^{(0)}(3,\,1')\tag{8.102}$$

So we can get G_2 and then G

- Too complicated!
- Hard to compute when n is large.

There is a simple way to directly get the finally answer we want, thanks the very smart technique designed by Feynman.

8.3.7. Feynman diagrams and Feynman rules

- Each integration coordinate is represented by a point;
- A propagator, G(1,1'), is represented by a solid line;
- A creation operator is represented by a solid line attached to the point with an arrow from the point;
- An annihilation operator is represented by a solid line attached to the point with an arrow to the point;
- The interaction $V(1-2)$ is represented by a dashed line connecting two points (r_1 and r_2) and each ending point has a creation operator and an annihilation operator.
- Connect the lines, and keep the direction of the arrows

Find all the diagrams obey the rules described above, each of them represent a term in the power series expansion of G.

8.3.8. k,ω-space

Assuming the system have momentum and energy conservation laws. In the k,ω-space, we use the same diagrams and in addition, we also:

- Assign momentum and frequency to each line and keep the momentum and energy conservation law at each ending point or crossing points.
- For each loop, there is a pair of unknown q and $i\,\Omega_n$ and we need to integrate/sum over them.
- For each fermionic loop, we get an extra factor of (-1), which comes the commutation relation.

$$G(k,\,i\omega_n) = G^{(0)}(k,\,i\omega_n) \pm i \int d\!q \sum_{i\,\Omega_n} G^{(0)}(k,\,i\omega_n)\, G^{(0)}(k,\,i\omega_n)\, V(0,\,0)\, G^{(0)}(k,\,i\,\Omega_n) +$$

$$i \int d\!q \sum_{i\,\Omega_n} G^{(0)}(k,\,i\omega_n)\, G^{(0)}(k,\,i\omega_n)\, G^{(0)}(k+q,\,i\omega_n + i\,\Omega_n)\, V(q,\,i\,\Omega_n) + ... \tag{8.103}$$

8.4. the Dyson's equation

8.4.1. The sum of a geometric series

How do we compute the sum of a geometric series:

$$X = a + a\,q + a\,q^2 + a\,q^3 + ... \tag{8.104}$$

First, we notice that

$$q\,X = a\,q + a\,q^2 + a\,q^3 + a\,q^4 + ... \tag{8.105}$$

then, we rewrite X as

$$X = a + a\,q + a\,q^2 + a\,q^3 + ... = a + q\,X \tag{8.106}$$

So,

$$X - q\,X = a \tag{8.107}$$

$$X = \frac{a}{1-q} \tag{8.108}$$

We can use the same trick to sum many Feynman diagrams.

8.4.2. example: Hartree-Fock

Rewrite Hartree and Fork term as:

$$G(k,\,i\omega_n) = G^{(0)}(k,\,i\omega_n) + G^{(0)}(k,\,i\omega_n)\, \Sigma_{\mathrm{HF}}(k,\,i\omega_n)\, G^{(0)}(k,\,i\omega_n) \tag{8.109}$$

where Σ is called the self-energy:

$$\Sigma_{\mathrm{HF}}(k,\,i\omega_n) = \pm i \int d\!q \sum_{i\,\Omega_n} V(0,\,0)\, G^{(0)}(k,\,i\,\Omega_n) + i \int d\!q \sum_{i\,\Omega_n} G^{(0)}(k+q,\,i\omega_n + i\,\Omega_n)\, V(q,\,i\,\Omega_n) \tag{8.110}$$

Now, we can include many higher order terms (we didn't include all the terms here, but we included a lot)

$$G(k, i\omega_n) = G^{(0)}(k, i\omega_n) + G^{(0)}(k, i\omega_n)\,\Sigma(k, i\omega_n)\,G^{(0)}(k, i\omega_n) + G^{(0)}(k, i\omega_n)\,\Sigma(k, i\omega_n)\,G^{(0)}(k, i\omega_n)\,\Sigma(k, i\omega_n)\,G^{(0)}(k, i\omega_n) +$$
$$G^{(0)}(k, i\omega_n)\,\Sigma(k, i\omega_n)\,G^{(0)}(k, i\omega_n)\,\Sigma(k, i\omega_n)\,G^{(0)}(k, i\omega_n)\,\Sigma(k, i\omega_n)\,G^{(0)}(k, i\omega_n) + \dots \tag{8.111}$$

the sum of these diagram give us

$$G(k, i\omega_n) = \frac{G^{(0)}(k, i\omega_n)}{1 - \Sigma_{\mathrm{HF}}(k, i\omega_n)\,G^{(0)}(k, i\omega_n)} = \frac{1}{\frac{1}{G^{(0)}(k, i\omega_n)} - \Sigma_{\mathrm{HF}}(k, i\omega_n)} = \frac{1}{i\omega_n - \epsilon(k) - \Sigma_{\mathrm{HF}}(k, i\omega_n)} \tag{8.112}$$

We found that effectively, the interactions changes the energy of a particle from $\epsilon(k)$ into $\in (\mathrm{k}) - \Sigma_{\mathrm{HF}}(\mathrm{k},\ \mathrm{i}\omega_{\mathrm{n}})$. This is why it is called the self-energy correction.

8.4.3. higher order diagrams and One-particle irreducible diagrams

This trick can be easily generalized to higher order diagrams and works for any diagram (just make sure that double counting are avoided). By using more complicated self-energy,

$$\Sigma(k, i\omega_n) = \pm i \int dq \sum_{i\,\Omega_n} V(0, 0)\,G^{(0)}(k, i\,\Omega_n) + i \int dq \sum_{i\,\Omega_n} G^{(0)}(k + q, i\omega_n + i\,\Omega_n)\,V(q, i\,\Omega_n) + \dots \tag{8.113}$$

we can get more accurate $G(k, i\omega_n)$.

The rule to avoid double counting is easy. One just need to include 1-particle irreducible diagrams in the self-energy Σ

One-particle irreducible diagrams means that if we cut one internal diagram, the diagram is still connected.

8.5. Screening and the Random Phase approximation

It turns out that the Hartree-Fock approximation doesn't work well for metals (with $1/r$ e-e interactions). In fact, the free fermion theory works much better. Notice that the Hartree-Fock is the first order approximation, while the free-fermion theory is the zeroth order approximation, so this means that by using a "better" (more accurate) theory, we are getting further way from the truth. Why?

It turns out that one needs to consider more diagrams than just the first order Hartree-Fock terms to get the right answer. In fact, after the "necessary" diagrams are all included, one find that the correct theory is very similar to a free-fermion theory. We only need to adjust three parameters in the free-fermion theory to get the correct low-energy theory for interacting electrons.

8.5.1. Coulomb interaction in the momentum

Fourier transform for the Coulomb interaction

$$V(q) = \int dr\, V(r)\, e^{-i\,q\cdot r} = ? \tag{8.114}$$

$$V(r) = \frac{e^2}{r} = e\, U(r) \tag{8.115}$$

where $U(r) = \frac{e}{r}$ is the potential induced by a point charge e.

$$\nabla^2 U(r) = -4\,\pi\,e\,\delta(r) \tag{8.116}$$

In q-space

$$-q^2\, U(q) = -4\,\pi\,e \tag{8.117}$$

$$U(q) = 4\,\pi\,\frac{e}{q^2} \tag{8.118}$$

$$V(q) = \frac{4\,\pi\,e^2}{q^2} \tag{8.119}$$

$V(q)$ is a singular function at $q \to 0$.

8.5.2. Permittivity and the Dyson's equation for interaction $V(q, i\,\omega_n)$

Notice that the interaction is also a line in Feynman diagrams, similar to G. So we can also treat V as a propagator similar as G.

It is worthwhile to emphasize here that this is just another way to group and organize the diagrams. We are not adding any extra diagrams. Neither did we change any physics here.

To avoid confusion, let's call the original interaction $V^{(0)}(q, i\,\omega_n) = 4\,\pi\,e^2/q^2$.

$$V(q, i\,\omega_n) = V^{(0)}(q, i\,\omega_n) \pm V^{(0)}(q, i\,\omega_n)\,\Pi^{(0)}(q, i\,\omega_n)\,V^{(0)}(q, i\,\omega_n) + \dots \tag{8.120}$$

Here, the leading order correction is $\Pi^{(0)}$ is called a bubble. This name comes from the fact that the Feynman diagram for $\Pi^{(0)}$ looks like a bubble

$$\Pi^{(0)}(q, i\,\omega_n) = \int d\,k \sum_{i\Omega_n} G^{(0)}(k, i\Omega_n)\,G^{(0)}(k+q, i\Omega_n + i\,\omega_n) \tag{8.121}$$

The Dyson's equation

$$V(q, i\,\omega_n) = \frac{1}{V^{(0)}(q, i\,\omega_n)^{-1} \mp \Pi(q, i\,\omega_n)} \tag{8.122}$$

Physical meaning of this Dyson's equation:

$$V(q, i\,\omega_n) = \frac{V^{(0)}(q, i\,\omega_n)}{1 \mp \Pi(q, i\,\omega_n)\,V^{(0)}(q, i\,\omega_n)} = \frac{V^{(0)}(q, i\,\omega_n)}{\epsilon(q, i\,\omega_n)} \tag{8.123}$$

Inside a material, the e-e interaction is different from the Coulomb's interaction in vacuum. This is described by the "permittivity" ϵ. The vacuum permittivity is ϵ_0, but in a material it is different. This Dyson's compute the permittivity.

In insulators, ϵ is a number in the limit $q \to 0$ and $\omega \to 0$. But in a metal,

$$\text{Lim}_{q\to 0, \omega\to 0}\,\epsilon(q, \omega) = \infty \tag{8.124}$$

So the permittivity changes the interaction dramatically.

8.5.3. Screening

Q: What is the E field in a perfect metal? (consider only the static E field)

A: $E = 0$

Ohm's law

$$j = \sigma\,E \tag{8.125}$$

For a perfect metal, $\sigma \to \infty$, so if we want to keep j finite, $E = 0$.

Another way to see this is to notice that electrons in a perfect metal can move freely. If we put a test charge inside a perfect metal, the electrons will redistribute and "screen" the charge perfectly.

So the Coulomb's law in a metal is:

$$V(q, \omega = 0) = 0 = \frac{V^{(0)}(q, \omega = 0)}{\epsilon(q, \omega = 0)} \tag{8.126}$$

So, $\epsilon(q, \omega = 0) = \infty$

In a real metal, the "screening" is not prefect, so $\epsilon(q, \omega = 0)$ is not always ∞, but $\epsilon\,(q = 0, \omega = 0) = \infty$. The potential $V(q, \omega = 0)$ is not always 0, but $V(r, \omega = 0)$ decays exponentially as a function of r (instead of $1/r$). In other words, in a metal the e-e interactions are short-range interactions (not $1/r$).

8.5.4. The random phase approximation

The self-energy

$$\Pi(q, i\,\omega_n) = \Pi^{(0)}(q, i\,\omega_n) + O(V^1) \tag{8.127}$$

If we only keep the leading order term $\Pi^{(0)}(q, i\,\omega_n)$, this is known as the RPA approximation

$$V^{\text{RPA}}(q, i\,\omega_n) = \frac{V^{(0)}(q, i\,\omega_n)}{\epsilon^{\text{RPA}}(q, i\,\omega_n)} \tag{8.128}$$

where

$$\epsilon^{\text{RPA}}(q, i\,\omega_n) = 1 + \Pi^{(0)}(q, i\,\omega_n) \, V^{(0)}(q, i\,\omega_n) \tag{8.129}$$

Then, we use V^{RPA} to substitute V and compute the Hartree-Fock self-energy Σ^{RPA}.

At $T = 0$,

$$\Pi^{(0)}(q, i\,\omega_n) = \text{constant} + O(q, i\,\omega_n) \tag{8.130}$$

And this constant is proportional to the density of state N(0).

For insulators, the DOS=0

$$\epsilon^{\text{RPA}}(q = 0, z = 0) = 1 \tag{8.131}$$

(Here, we are using a simplified model. We didn't consider the contributions from lattice sites etc. If those effect are included, $\epsilon(q = 0, \omega = 0)$ shall not be 1, but it is also a constant)

For metals, the DOS>0

$$\epsilon^{\text{RPA}}(q, z = 0) = 1 + \frac{\text{constant}}{q^2} \, 4\,\pi\,e^2 \tag{8.132}$$

$$\text{Lim}_{q \to 0} \, \epsilon^{\text{RPA}}(q, z = 0) = \infty \tag{8.133}$$

$$V^{\text{RPA}}(q, i\,\omega_n = 0) = \frac{4\,\pi\,e^2}{q^2 + \text{constant}\, 4\,\pi\,e^2} \tag{8.134}$$

Transfer back to the real space

$$V^{\text{RPA}}(r, i\,\omega_n = 0) = A \exp(-r/r_0) \tag{8.135}$$

The interaction decays exponentially as r increases.

For a Fermi liquid, the RPA shows pretty good agreement with experiments and is the minimum approximation to describe an interacting Fermi liquid.

$$G(k, i\omega_n) = \frac{1}{i\omega_n - \epsilon(k) - \Sigma^{\text{RPA}}(k, i\omega_n)} \tag{8.136}$$

$$G(k, z) = \frac{1}{z - \epsilon(k) - \Sigma^{\text{RPA}}(k, z)} \tag{8.137}$$

For a Fermi liquid at T=0 K, we focus on $k \sim k_F$ and z~0.

Near k_F, $\epsilon(k) = \mu + v(k - k_F) + O(k - k_F)^2$.

$$G^{(0)}(k_F, z) = \frac{1}{z - v(k - k_F) - \mu + O(k - k_F)^2} \tag{8.138}$$

Using this $G^{(0)}$, we can compute Σ^{RPA}

For $\Sigma^{\text{RPA}}(k, z)$, we expand it around $k \sim k_F$ and z~0.

$$G(k_F, z) = \frac{1}{z - v(k - k_F) - \mu - \Sigma^{\text{RPA}}(k_F, 0) - a(k - k_F) + b\,z + O(k - k_F)^2 \ + O(z)^2} \tag{8.139}$$

It is important to notice that b is real and positive (z^2 term has a complex coefficient).

$$G^{\text{RPA}}(k_F, z) = \frac{Z}{z - \tilde{v}(k - k_F) - \tilde{\mu}} + O\,(k - k_F)^2 \ + O(z)^2 \tag{8.140}$$

here $Z = 1/(1 + b) < 1$. $\tilde{v} = \frac{v + a}{1 + b}$ and $\tilde{\mu} = \frac{\mu + \Sigma^{\text{RPA}}(k_F, 0)}{1 + b}$.

Notice that $G^{\text{RPA}}(k_F, z)$ looks just like the Green's function for a free fermion, with a renoramlized Fermi velocity and a renormalized chemical potential. This is why we can treat electrons as "free" particles in a Fermi liquid, although the interaction is pretty strong.

In addition, we have an extra factor Z, which is known as the Fermi residue

$$G^{\mathrm{RPA}}(k_F,\, z) = Z\, G^{\mathrm{Free}}(k_F,\, z) \tag{8.141}$$

This extra factor is $0<Z<1$. The physical meaning is that now at k_F the jump in occupation number $n(k_F - \delta) - n(k_F + \delta) = Z$ not 1.

9

Quantum phase transition and quantum criticality

9.1. Basic Ideas

Classical phase transition: phase transition at finite temperature

Quantum phase transition: phase transition at $T = 0$.

9.1.1. Quantum phase transition and level crossing

Consider a quantum Hamiltonian:

$$H = (1 - x) H_1 + x H_2 \tag{9.1}$$

where x varies between 0 and 1. At $x = 0$, $H = H_1$. So the ground state at $x = 0$ is the ground state of H_1 ($|G_1\rangle$). At $x = 1$ $H = H_2$, so the ground state of the system is ground state of H_2 ($|G_2\rangle$).

Q: At x increases from 0 to 1, how does the ground state wavefunction of H evolves from $|G_1\rangle$ to $|G_2\rangle$?

A: there are two possibilities. (1) $|G_1\rangle$ evolves adiabatically to $|G_2\rangle$ or (2) there is a level crossing point at some critical $x = x_C$.

The second case is known as a quantum phase transition.

Q: Why the level crossing? (Why not avoided level crossing)

A: If the symmetry properties of $|G_1\rangle$ and $|G_2\rangle$ are fundamentally different,

9.1.2. Example: transverse field Ising model

$$H = -J g \sum_i \sigma_i^x - J \sum_{\langle i,j \rangle} \sigma_i^z \sigma_j^z \tag{9.2}$$

The first term $-J g \sum_i \sigma_i^x$ is our H_1 and the second term $-J \sum_{\langle i,j \rangle} \sigma_i^z \sigma_j^z$ is our H_2.

Two special limits are easy to study:

At $g = +\infty$, the first term dominates so $H = H_1$. Now, the ground state is: all spins point to the $+x$ direction, which is our $|G_1\rangle$. At very large g (but not infinite g), most of the spin will be pointing in the $+x$ direction.

At $g = 0$, the first term vanishes so $H = H_2$. Now, there two degenerate ground states: all spins point to the $+z$ or -z direction, which we call $|G_2^+\rangle$ and $|G_2^-\rangle$. At $0 < g << 1$, not all the spins will be pointing into the same direction, but we will still have two ground states, one has most of the spin pointing to $+z$ the other have most spin pointing to $-z$.

Symmetry properties

The Hamiltonian preserves the symmetry: $z \to -z$. For the ground state $|G_1\rangle$, the symmetry is preserved. But for $|G_2^\pm\rangle$, this symmetry is broken spontaneously. If we change z→-z, $|G_2^+\rangle \leftrightarrow |G_2^-\rangle$.

G_1 has one ground state which preserves the $z \to -z$ symmetry. $|G_2^\pm\rangle$ have two ground states, which break the $z \to -z$ symmetry.

So, it is impossible for G_1 to evolve adiabatically to G_2 as g decreases from $+\infty$ to 0. So there must be (at least) one phase transition between $g = 0$ and $g = +\infty$.

9.1.3. Quantum critical point and quantum criticality

For some quantum phase transitions, the transition point shows scaling behavior. This type of quantum phase transitions are known as "second order" quantum phase transitions and the transition points for these transitions are known as quantum critical points.

Correlation functions:

$$\langle \sigma_i^z \sigma_j^z \rangle \propto \exp(|i-j| a/\xi) \tag{9.3}$$

At long distance $|i-j| >> 1$, this correlation function is typically an exponential function. Here ξ is known as the correlation length and a is the lattice spacing. Near a quantum critical point,

$$\xi \propto |g - g_c|^{-\nu} \tag{9.4}$$

these type of power-law behavior is known as scaling law.

The energy gap near a quantum critical point also obey the scaling law

$$\Delta \propto |g - g_c|^{z\nu} = \xi^{-z} \tag{9.5}$$

Notice that $\hbar/\Delta$ defines a time scale t

$$t = \hbar/\Delta \propto |g - g_c|^{z\nu} = \xi^z \tag{9.6}$$

So

$$t \propto \xi^z \tag{9.7}$$

Time counts as z space dimension. This z is known as dynamic critical exponent. In fact, one can usually prove that for a system in d space dimensions, a quantum critical point with dynamic critical exponent z can be mapped to a classical phase transition in $d + z$ space dimensions with the same symmetry breaking pattern.

For example, the transverse Ising model has $z = 1$. Therefore, a d-dimensional transverse Ising model is Equivalent to a d+1 dimensional classical Ising model.

9.2. Quantum Ising model in 1D

$$H = -J g \sum_i \sigma_i^x - J \sum_{\langle i,j \rangle} \sigma_i^z \sigma_j^z \tag{9.8}$$

First let's rotate the system around y-axis by 90 degrees (changing coordinates, which doesn't effect any physics). $x \to -z$ and $z \to x$

$$H = J g \sum_i \sigma_i^z - J \sum_i \sigma_i^x \sigma_{i+1}^x \tag{9.9}$$

9.2.1. Jordan-Wigner transformation

$$\sigma_i^+ = (-1)^{\sum_{j=1}^{j=i-1} n_j} c_i^\dagger = \Pi_{j=1}^{j=i-1} (-1)^{n_j} c_i^\dagger = \Pi_{j=1}^{j=i-1} (1 - 2 n_j) c_i^\dagger \tag{9.10}$$

$$\sigma_i^- = (-1)^{\sum_{j=1}^{j=i-1} n_j} c_i \Pi_{j=1}^{j=i-1} (-1)^{n_j} c_i = \Pi_{j=1}^{j=i-1} (1 - 2 n_j) c_i \tag{9.11}$$

$$\sigma_i^z = 2 S^z = 2 (n_i - 1/2) = 2 n_i - 1 \tag{9.12}$$

$$\sigma_i^x = \sigma_i^+ + \sigma_i^- = \Pi_{j=1}^{j=i-1} (1 - 2 n_j)(c_i^\dagger + c_i) \tag{9.13}$$

$$\begin{aligned}
H = J g \sum_i \sigma_i^z - J \sum_i \sigma_i^x \sigma_{i+1}^x &= J g \sum_i (2 n_i - 1) - J \sum_i \Pi_{j=1}^{j=i-1}(1 - 2 n_j)(c_i^\dagger + c_i) \Pi_{j=1}^{j=i}(1 - 2 n_j)(c_{i+1}^\dagger + c_{i+1}) = \\
& J g \sum_i (2 c_i^\dagger c_i - 1) - J \sum_i \Pi_{j=1}^{j=i-1}(1 - 2 n_j)(c_i^\dagger + c_i) \Pi_{j'=1}^{j'=i-1}(1 - 2 n_{j'})(1 - 2 n_i)(c_{i+1}^\dagger + c_{i+1}) = \\
& J g \sum_i (2 c_i^\dagger c_i - 1) - J \sum_i \Pi_{j=1}^{j=i-1} \Pi_{j'=1}^{j'=i-1}(1 - 2 n_j)(1 - 2 n_{j'})(c_i^\dagger + c_i)(1 - 2 n_i)(c_{i+1}^\dagger + c_{i+1}) = \\
& J g \sum_i (2 c_i^\dagger c_i - 1) - J \sum_i \Pi_{j=1}^{j=i-1}(1 - 2 n_j)^2 (c_i^\dagger - c_i)(c_{i+1}^\dagger + c_{i+1})
\end{aligned} \tag{9.14}$$

The last step we used the fact that $c_i^\dagger(1 - 2 n_i) = c_i^\dagger$ and $c_i(1 - 2 n_i) = -c_i$

$$c_i^\dagger (1 - 2 n_i) |0\rangle = c_i^\dagger |0\rangle = |1\rangle \tag{9.15}$$

$$c_i^\dagger (1 - 2 n_i) |1\rangle = -c_i^\dagger |1\rangle = 0 \tag{9.16}$$

$$c_i (1 - 2 n_i) |0\rangle = c_i |0\rangle = 0 \tag{9.17}$$

$$c_i (1 - 2 n_i) |1\rangle = -c_i |1\rangle = |0\rangle \tag{9.18}$$

Compare with

$$c_i^\dagger |0\rangle = |1\rangle \tag{9.19}$$

$$c_i^\dagger |1\rangle = 0 \tag{9.20}$$

$$c_i |0\rangle = 0 \tag{9.21}$$

$$-c_i |1\rangle = |0\rangle \tag{9.22}$$

We find that $c_i^\dagger (1 - 2 n_i) = c_i^\dagger$ and $c_i (1 - 2 n_i) = -c_i$

$$H = J g \sum_i \left(2 c_i^\dagger c_i - 1\right) - J \sum_i \Pi_{j=1}{}^{j=i-1} \left(1 - 2 n_j\right)^2 \left(c_i^\dagger - c_i\right)\left(c_{i+1}^\dagger + c_{i+1}\right) \tag{9.23}$$

Because $\left(1 - 2 n_j\right)^2 = 1$,

$$H = J g \sum_i \left(2 c_i^\dagger c_i - 1\right) - J \sum_i \left(c_i^\dagger - c_i\right)\left(c_{i+1}^\dagger + c_{i+1}\right) =$$
$$-J \sum_i \left(c_i^\dagger c_{i+1} + c_{i+1}^\dagger c_i\right) - J \sum_i \left(c_i^\dagger c_{i+1}^\dagger + c_{i+1} c_i\right) + 2 J g \sum_i c_i^\dagger c_i - J g N \tag{9.24}$$

The last term is a constant number, which can be ignored.

The first term describes the nearest-neighbor hoppings. The second term is a pairing term (cooper pairs in a superconductor). The third term is chemical potential $\mu = -2 J g$.

Go to the momentum space

$$c_k = \frac{1}{\sqrt{N}} \sum_j c_j e^{-i k r_j} \tag{9.25}$$

$$H = -J \sum_k \left(c_k^\dagger c_k e^{i k a} + c_k^\dagger c_k e^{-i k a}\right) - J \sum_i \left(c_{-k}^\dagger c_k^\dagger e^{-i k a} + c_{-k} c_k e^{-i k a}\right) + 2 J g \sum_k c_k^\dagger c_k =$$
$$J \sum_k \left[2 \left(g - \cos k a\right) c_k^\dagger c_k + i \sin k a \left(c_{-k}^\dagger c_k^\dagger + c_{-k} c_k\right)\right] = J \sum_k \left(c_k^\dagger \quad c_{-k}^\dagger \right) \begin{pmatrix} g - \cos k a & i \sin k a \\ -i \sin k a & -g + \cos k a \end{pmatrix} \begin{pmatrix} c_k \\ c_{-k} \end{pmatrix} \tag{9.26}$$

Eigenvalues of the 2×2 matrix gives the energy spectrum of a quasi $-$ particle. $\tag{9.27}$

$$E_\pm = \pm J \sqrt{(g - \cos k a)^2 + (\sin k a)^2} = \pm J \sqrt{1 + g^2 - 2 g \cos k a} \tag{9.28}$$

The gap between the two bands

$$\Delta(k) = E_+ - E_- = 2 J \sqrt{1 + g^2 - 2 g \cos k a} \tag{9.29}$$

Minimum of $\Delta(k)$ is reached at k=0.

$$\Delta(k) \geq \Delta = 2 J \sqrt{1 + g^2 - 2 g} = 2 J \, |1 - g| \tag{9.30}$$

Δ is the gap between the ground state and exacted states. As one can see that $\Delta=0$ at g=1. So $g = 1$ is our quantum phase transition point. Near this quantum phase transition point, Δ shows critical scaling $\Delta \propto |g - g_c|$. So we have $z \nu = 1$.

Similarly, one can prove that $\xi \propto 1 / |1 - g|$ so $\nu=1$. So we have $z = 1$ and $\nu=1$.

9.2.2. degenerate ground states at g<1.

For g<1, we know that there should be two degenerate ground states. How can we see it in the fermion model?

As shown in the homework, this 1D superconductor is a topological superconductor for $-1 < g < 1$. And it is a topologically trivial superconduc-

tor for $|g| > 1$. The topological superconductor has two Majorana fermions at the two ends of the 1D chain. In addition, we also proved that if we have two Majorana fermions, the ground state is 2-fold degenerate. This is the two-fold degeneracy for the symmetry broken phase.