# Contents

# List of Figures

# Chapter 1

# Preliminaries

## 1.1 Introduction

The quantum Hall effect (QHE) refers to a set of phenomena and associated phases of matter found in two-dimensional electron gases subjected to a large perpendicular magnetic field[1]. The phenomena are typically divided into two classes, the *integer quantum Hall effect* (IQHE) and the *fractional quantum Hall effect* (FQHE), depending on the *Landau level filling fraction*, given by $\nu = nhc/eB$, where $n$ is the two-dimensional electron density and $B$ the magnetic field strength. The combination $\phi_0 = hc/e = 4.137 \times 10^{-7}\,\mathrm{G}\cdot\mathrm{cm}^2$ is the Dirac flux quantum[2], hence

$$\nu = 4.14 \cdot n[10^{11}\,\mathrm{cm}^{-2}]\big/B[\mathrm{T}] \quad . \tag{1.1}$$

Thus, in a field of $B = 4.14\,\mathrm{T}$, the Landau level (LL) filling fraction $\nu = 1$ occurs for an electron density $n = 10^{11}\,\mathrm{cm}^{-2}$.

The IQHE was discovered by von Klitzing in 1980 in routine magnetotransport studies of silicon MOSFETs[3]. The FQHE was discovered by Tsui and Störmer in 1982 [4], in GaAs$-$Al$_x$Ga$_{1-x}$As heterojunctions. The experimental setup is depicted in Fig. 1.1, and some spectacular data shown in Fig. 1.2. An electrical current $I$ is established along the $\hat{\boldsymbol{x}}$ direction, and the longitudinal and transverse voltage drops $V_{\mathrm{L}}$ and $V_{\mathrm{H}}$ are measured, from which one obtains, in the linear response regime, the resistances $R_{\mathrm{L}} = V_{\mathrm{L}}/I$ and $R_{\mathrm{H}} = V_{\mathrm{H}}/I$. In the IQHE, one observes that $R_{\mathrm{H}}$ remains constant along plateaus as the filling fraction $\nu$ is varied (either by varying the electron density $n$, typically with a gate, or by varying the magnetic field $B$). The plateau values are given by $R_{\mathrm{H}} = h/pe^2$, where $p \in \mathbb{Z}$ is an integer, for $\nu = p$. In the FQHE, one observes

[1] The effect has been seen in hole gases as well.

[2] This is often more conveniently expressed as $\phi_0 = 4.137 \times 10^5\,\mathrm{T}\cdot\mathrm{\AA}^2$, where $1\,\mathrm{T} = 10^4\,\mathrm{G}$.

[3] K. von Klitzing, G. Dorda, and M. Pepper, *Phys. Rev. Lett.* **45**, 494 (1980).

[4] D. C. Tsui, H. L. Störmer, and A. C. Gossard, *Phys. Rev. Lett.* **48**, 1559 (1982).

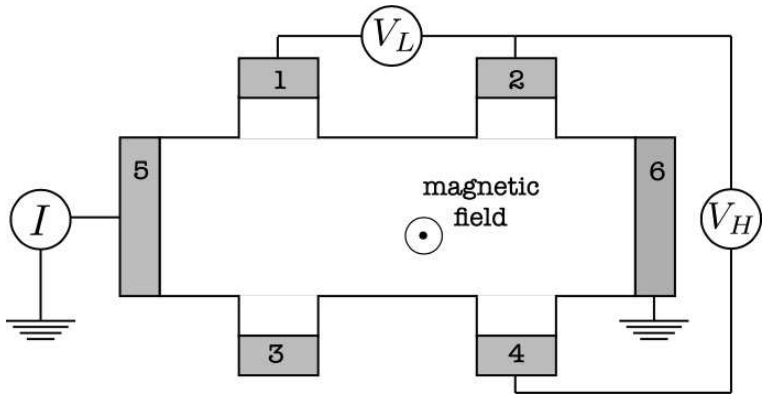


Figure 1.1: A Hall bar setup.

plateaus at rational fractions $\nu = p/q$, typically with $q$ odd[5], where $R_{\rm H} = qh/pe^2$. The quantity $R_{\rm Q} = h/e^2 = 25,812.8\,\Omega$ is known as the *quantum of resistance*.

### 1.1.1 Resistance, conductance, resistivity, conductivity

In the linear response regime, one has $V_\alpha = R_{\alpha\beta}\, I_\beta$, where $R$ is the resistance tensor. It's matrix inverse, $G = R^{-1}$, is known as the conductance, with $I_\alpha = G_{\alpha\beta}\, V_\beta$. The units of each element $R_{\alpha\beta}$ of the resistance tensor are Ohms ($\Omega$), hence the units of $G_{\alpha\beta}$ are $\Omega^{-1}$.

Resistance and conductance are not materials parameters (*i.e.* intensive quantities); you can't look up the resistance of copper in a table, for example. If, *ceteris paribus*, you double the length of a copper wire, its resistance doubles[6]. What doesn't change is the metal's *resistivity*, $\rho$, which *is* a materials parameter[7]. The corresponding linear response relation is between *current density* and *electric field*, *viz.* $E_\alpha = \rho_{\alpha\beta}\, j_\beta$. The inverse of the resistivity tensor is the conductivity tensor $\sigma = \rho^{-1}$, for which $j_\alpha = \sigma_{\alpha\beta} E_\beta$.

For an isotropic $d$-dimensional cube of side length $L$, in zero magnetic field, if the current along one of the cubic axes is $I$ then the current density is $j = I/L^{d-1}$. Similarly, if the voltage drop along this axis is $V$, the electric field is $E = V/L$. Thus $R = V/I = \rho L^{2-d}$, and we see that resistance and resistivity in general have different units. Similarly $G = \sigma L^{d-2}$. In two dimensions, resistance and resistivity have the same dimensions, but nevertheless resistance is a *geometric* quantity. Consider a $L_x \times L_y$ rectangular sample with conductivity tensor

$$\sigma = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{yx} & \sigma_{yy} \end{pmatrix} \quad , \tag{1.2}$$

[5]The even denominator quantum Hall effect is very interesting and distinct from the odd denominator effect.

[6]Assuming, that is, that the length $L$ is longer than the inelastic scattering (or *phase breaking*) length, $\ell_\phi$. For $L < \ell_\phi$, quantum interference effects become important and Ohm's law is no longer valid.

[7]The resistivity will in general depend on the temperature, and on the density and type of impurities present, as well as on the material itself.

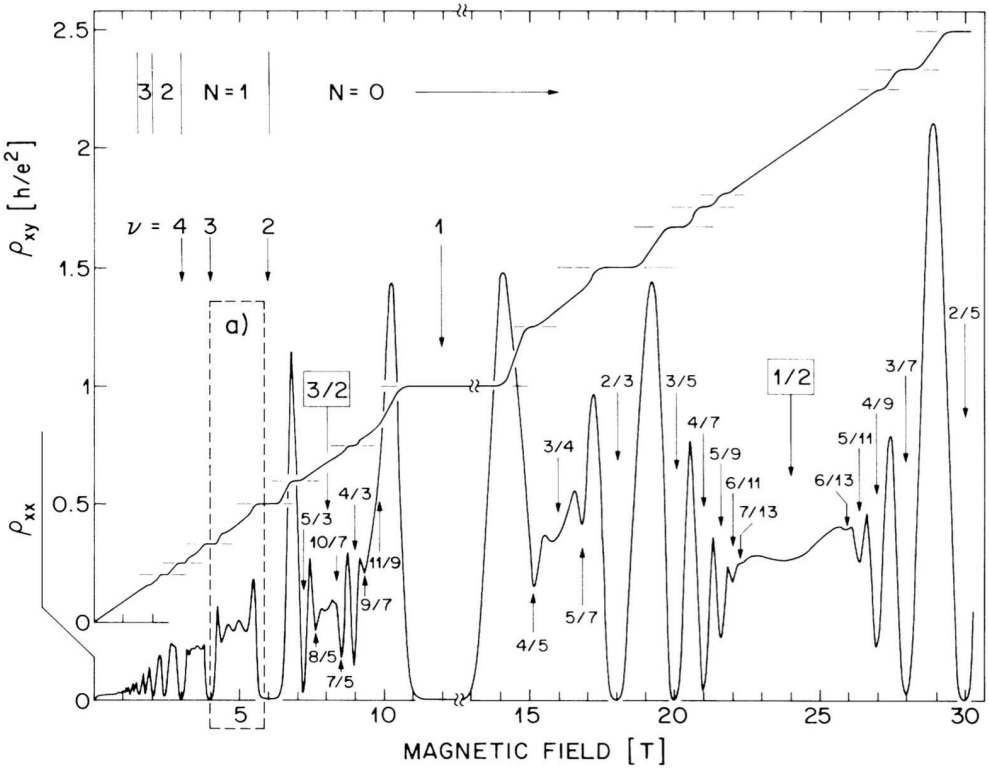


Figure 1.2: Low temperature ($T \approx 150\,\mathrm{mK}$) longitudinal resistivity $\rho_{xx}$ and Hall resistivity $\rho_{xy}$ as a function of applied magnetic field in a two-dimensional electron gas system (GaAs/AlGaAs heterostructure), from R. Willett *et al.*, *Phys. Rev. Lett.* **59**, 1776 (1987). Each dip in $\rho_{xx}$ and concomitant plateau in $\rho_{xy}$ corresponds to a distinct phase of matter.

with $\boldsymbol{j} = \sigma \boldsymbol{E}$. In general, linear response transport is described by the set of equations $J_i = L_{ik} F_k$, where the $\{J_i\}$ are generalized currents and the $\{F_k\}$ generalized forces. Onsager reciprocity[8] then requires

$$L_{ik}(\boldsymbol{B}) = \eta_i \, \eta_k \, L_{ki}(-\boldsymbol{B}) \quad , \tag{1.3}$$

with no sum on $i$ or $k$, where $\eta_i = \pm 1$ according to whether $J_i$ is symmetric or antisymmetric under time reversal, *i.e.* $J_i^{\mathcal{T}} = \eta_i J_i$. Thus, $\sigma_{yx}(\boldsymbol{B}) = \sigma_{xy}(-\boldsymbol{B})$ since both $j_x$ and $j_y$ are odd under time reversal. But $\boldsymbol{B} \to -\boldsymbol{B}$ reverses the orientation of the $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{B}})$ triad, hence $\sigma_{xy}(-\boldsymbol{B}) = -\sigma_{xy}(\boldsymbol{B})$, and we have that the off-diagonal elements of the conductivity tensor are antisymmetric: $\sigma_{yx}(\boldsymbol{B}) = -\sigma_{xy}(\boldsymbol{B})$. Now let's write the current densities as $j_x = I_x / L_y$ and $j_y = I_y / L_x$ , and the fields as $E_x = V_x / L_x$ and $E_y = V_y / L_y$ . We then have

$$\overbrace{\begin{pmatrix} L_y^{-1} & 0 \\ 0 & L_x^{-1} \end{pmatrix} \begin{pmatrix} I_x \\ I_y \end{pmatrix}}^{\boldsymbol{j}} = \overbrace{\begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ -\sigma_{xy} & \sigma_{yy} \end{pmatrix}}^{\sigma} \overbrace{\begin{pmatrix} L_x^{-1} & 0 \\ 0 & L_y^{-1} \end{pmatrix} \begin{pmatrix} V_x \\ V_y \end{pmatrix}}^{\boldsymbol{E}} \quad , \tag{1.4}$$

[8] See L. D. Landau and E. M. Lifshitz, *Statistical Physics*, part I, §120.

from which we read off the relation between conductance and conductivity tensors,

$$\begin{pmatrix} G_{xx} & G_{xy} \\ G_{yx} & G_{yy} \end{pmatrix} = \begin{pmatrix} L_y & 0 \\ 0 & L_x \end{pmatrix} \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ -\sigma_{xy} & \sigma_{yy} \end{pmatrix} \begin{pmatrix} L_x^{-1} & 0 \\ 0 & L_y^{-1} \end{pmatrix} = \begin{pmatrix} \frac{L_y}{L_x}\,\sigma_{xx} & \sigma_{xy} \\ -\sigma_{xy} & \frac{L_x}{L_y}\,\sigma_{yy} \end{pmatrix} \quad . \tag{1.5}$$

Similarly, the relation between resistance and resistivity tensors is

$$\begin{pmatrix} R_{xx} & R_{xy} \\ R_{yx} & R_{yy} \end{pmatrix} = \begin{pmatrix} \frac{L_x}{L_y}\,\rho_{xx} & \rho_{xy} \\ -\rho_{xy} & \frac{L_y}{L_x}\,\rho_{yy} \end{pmatrix} \quad . \tag{1.6}$$

Finally,

$$\rho = \begin{pmatrix} \rho_{xx} & \rho_{xy} \\ -\rho_{xy} & \rho_{yy} \end{pmatrix} = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ -\sigma_{xy} & \sigma_{yy} \end{pmatrix}^{-1} = \frac{1}{\sigma_{xx}^2 + \sigma_{xy}^2} \begin{pmatrix} \sigma_{yy} & -\sigma_{xy} \\ \sigma_{xy} & \sigma_{xx} \end{pmatrix} = \sigma^{-1} \quad . \tag{1.7}$$

Along the QH plateaus, as $T \to 0$, the longitudinal resistivity vanishes as $\rho_{xx}(T) \propto e^{-\Delta/k_{\rm B}T}$, where $\Delta$ is the energy gap for transport. Thus, at $T = 0$ the resistivity and conductivity tensors are purely off-diagonal, with $\rho_{\alpha\beta} = \rho_{xy}\,\epsilon_{\alpha\beta}$ and $\sigma_{\alpha\beta} = \sigma_{xy}\,\epsilon_{\alpha\beta}$ , with $\rho_{xy} = 1/\sigma_{xy}$.

### 1.1.2 Semiclassical magnetotransport theory

Combining Newton's second law with the Lorentz force law for a particle of charge $-e$ and mass $m$, we have

$$\frac{d\boldsymbol{p}}{dt} = -e\boldsymbol{E} - \frac{e}{c}\,\frac{\boldsymbol{p}}{m} \times \boldsymbol{B} - \frac{\boldsymbol{p}}{\tau} \quad , \tag{1.8}$$

where the last term is a frictional force which in metals and semiconductors typically comes from electron-impurity scattering[9], with $\tau$ the transport scattering time[10]. We take $\boldsymbol{B} = B\hat{\boldsymbol{z}}$, and write the current density as $\boldsymbol{j} = -ne\boldsymbol{p}/m$. Defining the *cyclotron frequency* $\omega_{\rm c} = eB/mc$, and setting $\dot{\boldsymbol{p}} = 0$ in steady state, we obtain

$$\frac{ne^2\tau}{m}\,\boldsymbol{E} + \omega_{\rm c}\tau\,\boldsymbol{j} \times \hat{\boldsymbol{z}} + \boldsymbol{j} = 0 \quad , \tag{1.9}$$

the solution of which is $\boldsymbol{j} = \sigma\boldsymbol{E}$, where the conductivity tensor is

$$\sigma = \frac{ne^2\tau/m}{1 + \omega_{\rm c}^2\tau^2} \begin{pmatrix} 1 & -\omega_{\rm c}\tau \\ \omega_{\rm c}\tau & 1 \end{pmatrix} \quad . \tag{1.10}$$

Taking the inverse, we have $\boldsymbol{E} = \rho\boldsymbol{j}$, with resistivity tensor

$$\rho = \sigma^{-1} = \frac{m}{ne^2\tau} \begin{pmatrix} 1 & \omega_{\rm c}\tau \\ -\omega_{\rm c}\tau & 1 \end{pmatrix} \quad . \tag{1.11}$$

[9] Electron-phonon scattering, electron-electron scattering, and boundary scattering are also present.

[10] There is an important difference between the single particle scattering time $\tau_{\rm sp}$ and the transport scattering time $\tau_{\rm tr}$. See, *e.g.*, §1.5 of my Physics 211B lecture notes for details.

What is $n$? Naïvely one might think it is the total electron density, but of course this is wrong. As we know from elementary solid state physics, filled Bloch bands are inert and carry no net current. A somewhat more realistic linearized Boltzmann equation approach, assuming an isotropic parabolic conduction band with electron carriers, yields the same result, with $n = \int d\varepsilon\, g_{\rm c}(\varepsilon)\, f^0(\varepsilon - \mu)$ the conduction electron density, with $g_{\rm c}(\varepsilon)$ is the conduction band density of states and $f^0(\varepsilon - \mu)$ the Fermi function, and $m$ replaced by the effective mass $m^*$ of the conduction band. All the fully occupied bands below the conduction band contribute nothing to the current. Note that $\rho_{xx} = \rho_{yy} = m^*/ne^2\tau$ because the system is isotropic. For the anisotropic parabolic band, where the effective mass tensor $m^*_{\alpha\beta}$ has eigenvalues $m^*_{x,y}$ , then along its principal axes one of course has $\rho_{xx} = m^*_x/ne^2\tau$ and $\rho_{yy} = m^*_y/ne^2\tau$, with $\rho_{xy} = B/nec$ as in the isotropic case.

One interesting feature of the semiclassical Boltzmann result is that the diagonal terms of the resistivity tensor are independent of magnetic field. Thus, $\partial\rho_{xx}/\partial B = 0$, and the *magnetoresistance* $\Delta\rho_{xx}(B) \equiv \rho_{xx}(B) - \rho_{xx}(0)$ vanishes. This is in general not the case if one has multiple bands contributing to the transport current (say conduction electrons as well as valence holes), or in the case where the Fermi surface has open orbits which span the Brillouin zone. Thus, as a function of $B$, the semiclassical result says that $\rho_{xx}(B)$ is constant and $\rho_{xy}(B)$ is perfectly linear. This it completely different from the results shown in Fig. 1.2, except in the very low field regime.

### 1.1.3 Mobility, cyclotron frequency, and electron-electron interactions

The *mobility* $\mu$ is defined by the combination $\mu = e\tau/m^*$. Thus, in zero field, the steady state electron velocity is $\boldsymbol{v} = \mu\boldsymbol{E}$, so mobility has units of $[\mu] = \mathrm{cm}^2/\mathrm{V}\cdot\mathrm{s}$. In MOSFETs, mobilities are seldom more than a few tens of thousands in these units. But in MBE-grown GaAs heterostructures, mobilities as high as $10^7\,\mathrm{cm}^2/\mathrm{V}\cdot\mathrm{s}$ have been achieved. In GaAs, where the conduction band is isotropic and has effective mass $m^* = 0.067\,m_{\rm e}$, one finds

$$\tau = 3.8 \times 10^{-17}\mathrm{s}\cdot\mu\big[\mathrm{cm}^2/\mathrm{V}\cdot\mathrm{s}\big] \quad . \tag{1.12}$$

Thus, for $\mu = 10^6\,\mathrm{cm}^2/\mathrm{V}\cdot\mathrm{s}$, one obtains $\tau \simeq 38\,\mathrm{ps}$.

The *cyclotron frequency* is given by the combination $\omega_{\rm c} = eB/m^*c$. With

$$\phi_0 = \frac{hc}{e} = 4.14{\times}10^{-7}\mathrm{G}\cdot\mathrm{cm}^2 \quad , \quad h = 6.63{\times}10^{-27}\mathrm{erg}\cdot\mathrm{s} = 4.14{\times}10^{-15}\mathrm{eV}\cdot\mathrm{s} \quad , \quad k_{\rm B} = 8.62{\times}10^{-5}\mathrm{eV/K} \quad , \tag{1.13}$$

Thus, for GaAs conduction electrons, one obtains

$$\omega_{\rm c} = 2.63{\times}10^{12}\,\mathrm{Hz}B[\mathrm{T}] \quad , \quad \omega_{\rm c}\tau = 10^{-4}\,\mu[\mathrm{cm}^2/\mathrm{Vs}]\,B[\mathrm{T}] \quad , \quad \hbar\omega_{\rm c} = 1.73\,\mathrm{meV}B[T] = 20\,\mathrm{K}\,k_{\rm B}\,B[T] \quad . \tag{1.14}$$

At fields $B \sim 10\,\mathrm{T}$ and in samples of mobility $\mu \sim 10^6\,\mathrm{cm}^2/\mathrm{V}\cdot\mathrm{s}$, we have $\omega_{\rm c}\tau \sim 1000 \gg 1$.

As we shall see, quantization introduces a new length scale, $\ell = (\hbar c/eB)^{1/2}$, called the *magnetic length*. This depends only on physical constants and the magnetic field strength. One finds

$$\ell = (\phi_0/2\pi B)^{1/2} = 257\,\text{Å}/\sqrt{B[T]} \quad . \tag{1.15}$$

From this length scale, we construct the energy scale $e^2/\epsilon\ell$ for electron-electron interactions. For GaAs, where $\epsilon = 13$, we have

$$\frac{e^2}{\epsilon\ell} = 4.31\,\text{meV}\cdot\sqrt{B[T]} = 50.0\,\text{K}\,k_{\text{B}}\sqrt{B[T]} \quad . \tag{1.16}$$

### 1.1.4 $\vec{E}\times\vec{B}$ drift and separation of time scales

For a classical particle of charge $e$ moving in the $(x,y)$ plane and subjected to a magnetic field $\boldsymbol{B} = B\hat{\boldsymbol{z}}$, the equations of motion are given by the Lorentz force law,

$$m\ddot{\boldsymbol{r}} = -\boldsymbol{\nabla}V - \frac{e}{c}\,B\,\dot{\boldsymbol{r}}\times\hat{\boldsymbol{z}} \quad . \tag{1.17}$$

We now write $\boldsymbol{r}(t) = \boldsymbol{\mathcal{R}}(t) + \boldsymbol{\xi}(t)$. We presume that the guiding-center motion $\boldsymbol{\mathcal{R}}(t)$ executes large excursions, slowly drifting along equipotentials of $V(\boldsymbol{r})$, while the cyclotron motion $\boldsymbol{\xi}(t)$ executes fast small excursions with characteristic time scale $2\pi/\omega_{\text{c}}$. This assumption will be borne out in the following analysis.

The zeroth order theory is simply given by

$$\begin{aligned}\dot{\boldsymbol{\mathcal{R}}} &= -\frac{c}{eB}\,\hat{\boldsymbol{z}}\times\boldsymbol{\nabla}V(\boldsymbol{\mathcal{R}}) \\ \dot{\boldsymbol{\xi}} &= \omega_{\text{c}}\,\hat{\boldsymbol{z}}\times\boldsymbol{\xi} \quad .\end{aligned} \tag{1.18}$$

Thus, the guiding-center executes a slow drift in the direction of $\boldsymbol{\nabla}V\times\hat{\boldsymbol{z}}$, while the cyclotron coordinate executes counterclockwise circular motion as viewed from above.

Proceeding with the expansion in powers of the cyclotron motion, we have

$$\begin{aligned}m\ddot{\mathcal{R}}_\alpha + m\ddot{\xi}_\alpha = &-\partial_\alpha V(\boldsymbol{\mathcal{R}}) - \xi_\beta\,\partial_\alpha\,\partial_\beta\,V(\boldsymbol{\mathcal{R}}) - \tfrac{1}{2}\,\xi_\beta\,\xi_\gamma\,\partial_\alpha\,\partial_\beta\,\partial_\gamma\,V(\boldsymbol{\mathcal{R}}) + \ldots \\ &- \frac{eB}{c}\,\epsilon_{\alpha\beta}\,\dot{\mathcal{R}}_\beta - \frac{eB}{c}\,\epsilon_{\alpha\beta}\,\dot{\xi}_\beta \quad .\end{aligned} \tag{1.19}$$

Here we have used the relation, for any vector $\boldsymbol{u}$,

$$\epsilon_{\alpha\beta}\,u_\beta = (u_y\,,\,-u_x) = (\boldsymbol{u}\times\hat{\boldsymbol{z}})_\alpha \quad . \tag{1.20}$$

We assume $\dot{\mathcal{R}}_\alpha = 0$ on average, leading to the slow equation,

$$\frac{eB}{c}\,\epsilon_{\alpha\beta}\,\dot{\mathcal{R}}_\beta = -\partial_\alpha V(\boldsymbol{\mathcal{R}}) - \tfrac{1}{2}\,\big\langle\xi_\beta\,\xi_\gamma\big\rangle\,\partial_\alpha\,\partial_\beta\,\partial_\gamma\,V(\boldsymbol{\mathcal{R}}) - \ldots \quad , \tag{1.21}$$

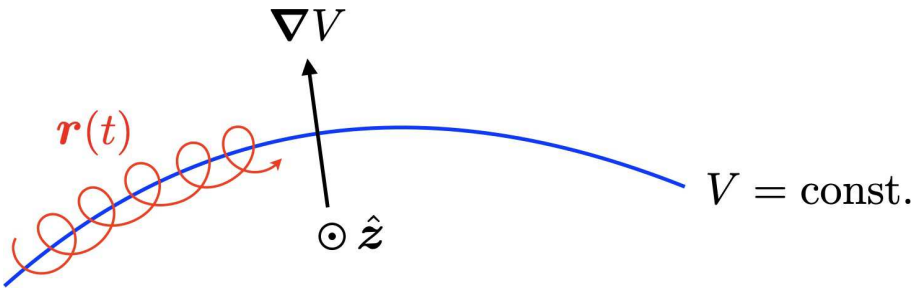


Figure 1.3: Cyclotron motion and guiding-center drift.

where $\langle \xi_\beta \, \xi_\gamma \rangle$ is averaged over the fast motion, and the fast equation,

$$m\ddot{\xi}_\alpha = -\xi_\beta \, \partial_\alpha \, \partial_\beta \, V(\boldsymbol{\mathcal{R}}) - \frac{eB}{c} \, \epsilon_{\alpha\beta} \, \dot{\xi}_\beta + \ldots \quad . \tag{1.22}$$

On the fast scale of the $\boldsymbol{\xi}(t)$ motion, the guiding-center $\boldsymbol{\mathcal{R}}(t)$ is assumed constant. Fourier transforming the fast motion, we write $\boldsymbol{\xi}(t) = \mathsf{Re}\, \boldsymbol{\xi}^0 \, e^{-i\omega t}$, with

$$\begin{pmatrix} -m\omega^2 + V_{xx} & im\omega\omega_{\rm c} + V_{xy} \\ -im\omega\omega_{\rm c} + V_{xy} & -m\omega^2 + V_{yy} \end{pmatrix} \begin{pmatrix} \xi^0_x \\ \xi^0_y \end{pmatrix} = 0 \quad , \tag{1.23}$$

where $V_{\alpha\beta} \equiv \partial_\alpha \, \partial_\beta \, V(\boldsymbol{\mathcal{R}})$. Solving for $\omega$, we take the fast root of the resulting quadratic equation and obtain

$$\begin{aligned} \omega_+^2 &= \frac{1}{2}\left(\omega_{\rm c}^2 + \frac{V_{xx}+V_{yy}}{m}\right) + \frac{1}{2}\,\omega_{\rm c}^2 \sqrt{1 + \frac{2(V_{xx}+V_{yy})}{m\omega_{\rm c}^2} + \frac{(V_{xx}-V_{yy})^2}{m^2\omega_{\rm c}^4} + \frac{V_{xy}^2}{m^2\omega_{\rm c}^4}} \\ &= \omega_{\rm c}^2 + \frac{\nabla^2 V}{m} + \ldots \quad . \end{aligned} \tag{1.24}$$

Thus the local cyclotron frequency is given by $\omega_{\rm c}(\boldsymbol{\mathcal{R}}) = \omega_{\rm c} + \frac{\nabla^2 V(\boldsymbol{\mathcal{R}})}{2m\omega_{\rm c}}$ to lowest nontrivial order.

We will need the corresponding eigenvector for the high frequency root. Writing $\boldsymbol{\xi}^0 \equiv (u\,\xi_0, v\,\xi_0)$, with $|u|^2 + |v|^2 = 1$, we have

$$\begin{aligned} u &= \frac{V_{xy} + im\omega_{\rm c}\omega_+}{\sqrt{(V_{xx} - m\omega_+^2)^2 + |V_{xy} + im\omega_{\rm c}\omega_+|^2}} \\ v &= -\frac{V_{xx} - m\omega_+^2}{\sqrt{(V_{xx} - m\omega_+^2)^2 + |V_{xy} + im\omega_{\rm c}\omega_+|^2}} \quad . \end{aligned} \tag{1.25}$$

Averaging over the cyclotron motion, we find

$$\langle \xi_\alpha \, \xi_\beta \rangle = \tfrac{1}{2}\,\xi_0^2 \begin{pmatrix} |u|^2 & \mathsf{Re}\,(u\bar{v}) \\ \mathsf{Re}\,(u\bar{v}) & |v|^2 \end{pmatrix} \quad . \tag{1.26}$$

Since $\omega_+ \approx \omega_{\rm c}$, we obtain $u \approx \frac{i}{\sqrt{2}}$ and $v \approx \frac{1}{\sqrt{2}}$. Thus the guiding-center motion is given by the equation

$$\frac{eB}{c}\,\epsilon_{\alpha\beta}\,\dot{\mathcal{R}}_\beta = -\partial_\alpha V - \tfrac{1}{4}\,\xi_0^2\,\partial_\alpha\,\nabla^2 V \equiv -\partial_\alpha V_{\rm eff}(\boldsymbol{\mathcal{R}}) \quad , \tag{1.27}$$

where the effective guiding-center potential is

$$V_{\rm eff}(\boldsymbol{\mathcal{R}}) = V(\boldsymbol{\mathcal{R}}) + \tfrac{1}{4}\,\xi_0^2\,\nabla^2 V(\boldsymbol{\mathcal{R}}) \quad . \tag{1.28}$$

This makes good physical sense: as the electron moves slowly along the equipotential, it samples, through its small and fast cyclotron excursions, the local environment, inducing a gradient squared correction to the local value of $V(\boldsymbol{\mathcal{R}})$.

For a classical electron moving in a circular orbit of radius $r$, setting the centrifugal force $F_{\rm c} = mv^2/r$ equal to the Lorentz force $evB/c$ yields the relation $v = eBR/mc$. The kinetic energy is then $T = \frac{1}{2}mv^2 = e^2B^2r^2/2mc^2$. If we now quantize semiclassically, demanding $\pi r^2 \cdot B = (n+\frac{1}{2})\phi_0$, then $r_n^2 = (2n+1)\ell^2$ where $\ell = (\hbar c/eB)^{1/2}$ is the magnetic length. The kinetic energy is then $T = (n+\frac{1}{2})\hbar\omega_{\rm c}$. Thus $\xi_0 = r_n$ in our above derivation of the effective potential, with $n$ the Landau level index.

The potential $V(\boldsymbol{r})$ is due to extrinsic disorder, arising typically from the irregular placement of the dopant atoms in a heterostructure, or semiconductor-oxide interface disorder in a MOSFET. In heterostructures, the dopant ions are typically several hundred Ångstroms removed from the 2DEG layer, and $V(\boldsymbol{r})$ is smooth on this length scale. Suppose the two dimensional electron gas lies in the plane $z = 0$ and consider a 'delta doping' profile in which the donor density is $N_{\rm d}(x,y,z) = N_{\rm d}(\boldsymbol{r})\,\delta(z-d)$ where $d$ is the distance between the 2DEG and the dopant layer. The electrical potential $\phi(\boldsymbol{r})$ at $\boldsymbol{r} = (x,y)$ in the 2DEG plane is then given by

$$\phi(\boldsymbol{r}) = \int \frac{d^2q}{(2\pi)^2} \int\limits_{-\infty}^{\infty} \frac{dq_z}{2\pi}\,\hat{N}_{\rm d}(\boldsymbol{q})\,e^{i\boldsymbol{q}\cdot\boldsymbol{r}}\,e^{iq_z d}\,\frac{4\pi e}{\boldsymbol{q}^2+q_z^2} = \int \frac{d^2q}{(2\pi)^2}\,\hat{N}_{\rm d}(\boldsymbol{q})\,e^{i\boldsymbol{q}\cdot\boldsymbol{r}}\,\frac{2\pi e\,\exp(-|\boldsymbol{q}|d)}{|\boldsymbol{q}|} \quad , \tag{1.29}$$

and we see that the components of $\hat{N}_{\rm d}(\boldsymbol{q})$ with high spatial frequency are attenuated exponentially. This smooths out the random potential experienced by the electrons in the 2DEG. MOSFETs are typically much dirtier, with correspondingly lower mobilities, hence $V(\boldsymbol{r})$ there is disorder on shorter length scales. Indeed disorder is *essential* to the quantum Hall effect, since in a pristine system we can always perform a Lorentz boost to a frame where $\boldsymbol{B} = 0$ and deduce $\sigma_{xy} = -nec/B$. (This argument is quite a bit more subtle if there are other features breaking translational symmetry, such as leads and surfaces.)

## 1.2 MOSFETs and Heterojunctions

Where do two-dimensional electron gases (2DEGs) come from? As noted above, the IQHE was first discovered in silicon MOSFETs while the FQHE was first discovered in GaAs heterostruc-

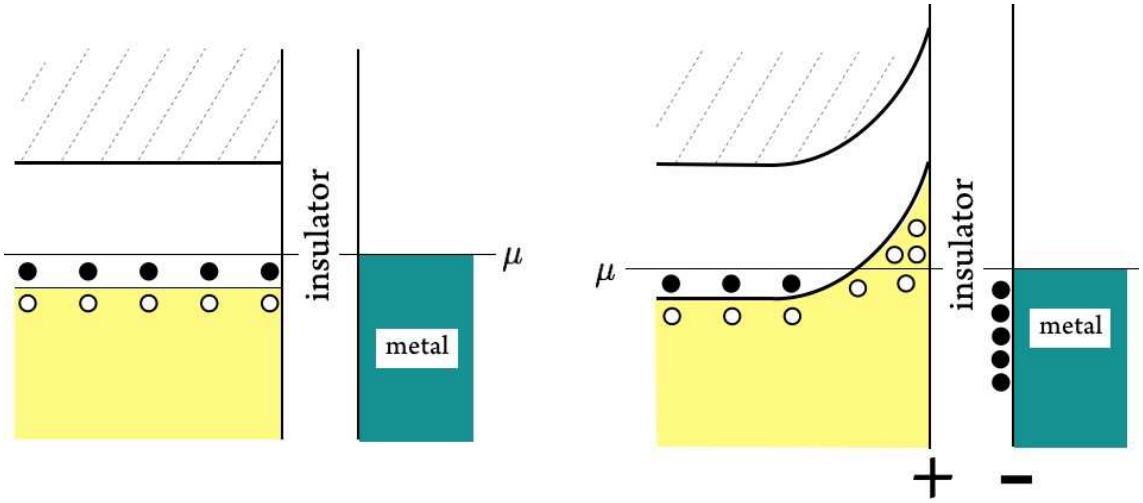


Figure 1.4: Junction between a $p$-type semiconductor and a metal. Left: Zero bias. Right: Metal biased negative with respect to semiconductor, creating an accumulation layer of holes and a net dipole moment at the interface.

tures. Details of the modeling and important semiconductor physics in these systems are discussed in the 1982 review by Ando, Fowler, and Stern[11]. Today, we have new two-dimensional systems which exhibit the QHE, such as graphene. Graphene is particularly interesting because it is a 'Dirac material' in which the electronic band structure features Dirac points, which are conical intersections of conduction and valence bands described by a two-dimensional Dirac Hamiltonian. More on this later.

In a metal, internal electric fields are efficiently screened and excess charge migrates rapidly to the surface, with charge density fluctuations attenuated exponentially as one enters the bulk. The Thomas-Fermi screening length, $\lambda_{\rm TF} = \big(4\pi e^2 g(\varepsilon_{\rm F})\big)^{-1/2}$, is short (a few Ångstroms) due to the large density of states at the Fermi level. In semiconductors, the Fermi level lies somewhere in the gap between valence and conduction bands, and the density of states at $\varepsilon_{\rm F}$ is quite low. Screening is due to thermally excited charge carriers, and since the carrier density is small in comparison to that in metals, the screening length is many lattice spacings.

Consider now a junction between a semiconductor and a metal, with an intervening insulating layer. This is called MIS, or Metal-Insulator-Semiconductor, junction. If the metal is unbiased relative to the semiconductor, their chemical potentials will align. The situation for a $p$-type semiconductor - metal junction is depicted in the left panel of Fig. 1.4. Next consider the case in which the metal is biased negatively with respect to the semiconductor, *i.e.* the metal is placed at a negative voltage $-V$. There is then an electric field $\boldsymbol{E} = -\boldsymbol{\nabla}\phi$ pointing *out* of the semiconductor. Electric fields point in the direction positive charges want to move, hence in this case valence holes are attracted to the interface, creating an *accumulation layer* of holes, as

[11] T. Ando, A. B. Fowler, and F. Stern, *Rev. Mod. Phys.* **54**, 437 (1982).

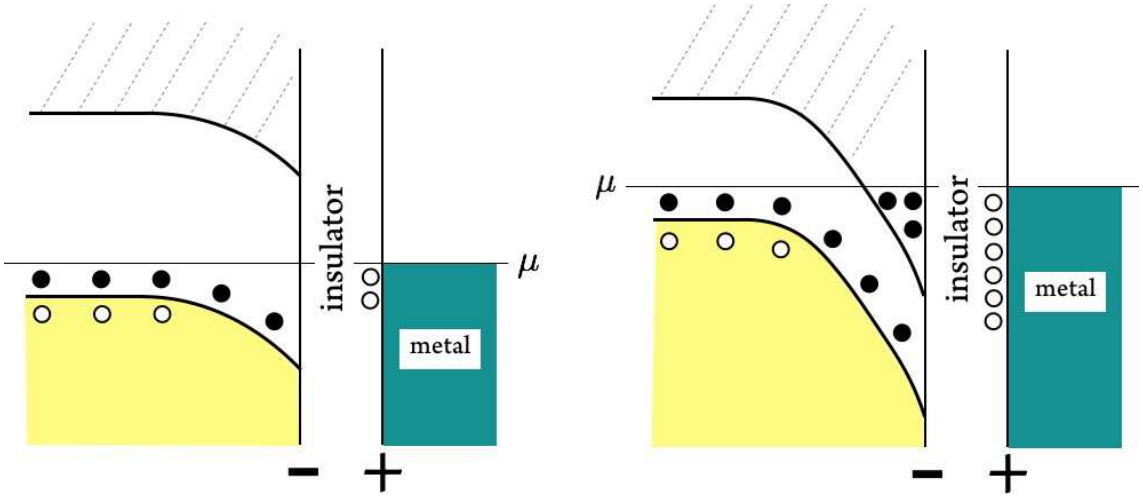


Figure 1.5: Junction between a $p$-type semiconductor and a metal. Left: Metal biased positive with respect to semiconductor, creating a space charge depletion layer. Right: Strong positive bias creates an inversion layer of $n$-type carriers in the $p$-type material.

depicted in the right panel of Fig. 1.4. On the metallic side, electrons migrate to the interface for the same reason. *No charges move across the insulating barrier.* Thus, a dipole layer is created across the barrier, with the dipole moment pointing into the semiconductor. This creates an internal potential whose net difference $\phi_{\rm metal} - \phi_{\rm semiconductor}(-\infty) = V$ exactly cancels the applied bias. This condition in fact is what determines the width of the accumulation layer.

What happens when the metal is biased positively? In this case, the electric field points into the semiconductor, and valence holes are repelled from the semiconductor surface, which is then negatively charged. This, in turn, repels electrons from the nearby metallic surface. The result is a space charge *depletion layer* in the semiconductor, which is devoid of charge carriers (*i.e.* valence holes). This situation is sketched in the left panel of Fig. 1.5.

Finally, what happens if the bias voltage on the metal exceeds the band gap? In this case, the field is so strong that not only are valence holes expelled from the surface, but conduction electrons are present, as shown in the right panel of Fig. 1.5. The presence of $n$-type carriers in a $p$-type semiconductor is known as *$n$-inversion*.

Remember this:

- *Accumulation* : presence of additional $n$-carriers in an $n$-type material, or additional $p$-carriers in a $p$-type material.
- *Depletion* : absence of $n$-carriers in an $n$-type material, or $p$-carriers in a $p$-type material.
- *Inversion* : presence of $n$-carriers in a $p$-type material, or $p$-carriers in an $n$-type material.

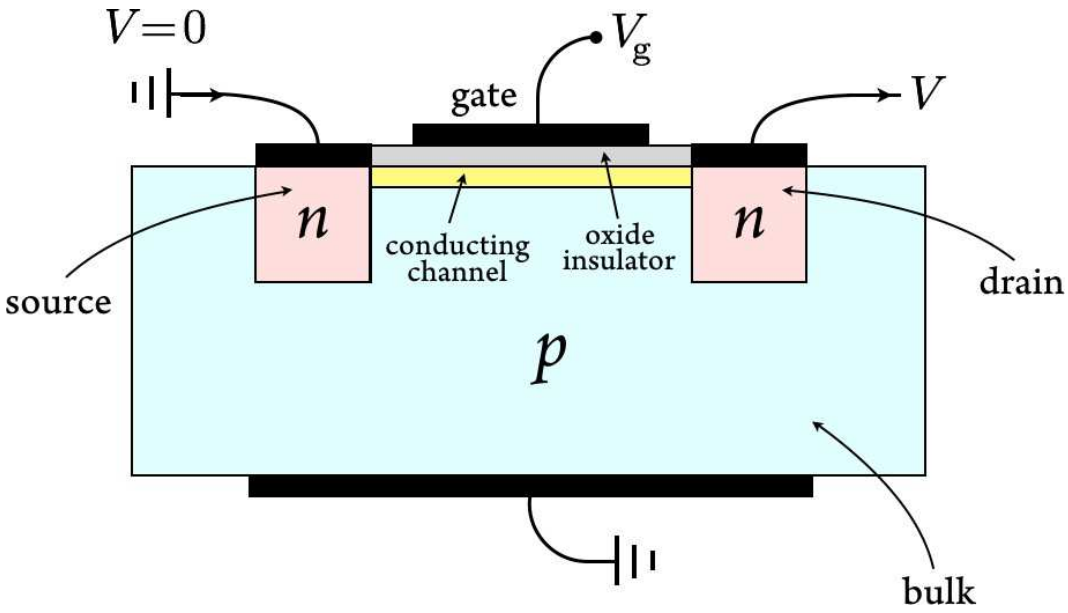


Figure 1.6: The MOSFET.

Inversion occurs when the presence of a depletion layer does not suffice to align the chemical potentials of the two sides of the junction.

### 1.2.1 The MOSFET

A MOSFET (Metal-Oxide-Semiconductor-Field-Effect-Transistor) consists of two back-to-back $p$-$n$ junctions, and, transverse to this, a gate-bulk-oxide capacitor. The situation is depicted in Fig. 1.6. If there is no gate voltage ($V_{\rm g} = 0$), then current will not flow at any bias voltage $V$ because necessarily one of the $p$-$n$ junction will be reverse-biased. The situation changes drastically if the gate is held at a high positive potential $V_{\rm g}$, for then an $n$-type accumulation layer forms at the bulk-gate interface, thereby connecting source and drain directly and resulting in a gate-controlled current flow. Although not shown in the figure, generally both source and drain are biased positively with respect to the bulk in order to avoid current leakage.

### 1.2.2 Heterojunctions

Potential uses of a junction formed from two distinct semiconductors were envisioned as early as 1951 by Shockley. Such devices, known as *heterojunctions*, have revolutionized the electronics industry and experimental solid state physics, the latter due to the advent of epitaxial technology which permits growth patterning to nearly atomic precision. Whereas the best inversion layer mobilities in Si MOSFETs are $\mu \approx 2 \times 10^4\,{\rm cm}^2/{\rm V\,s}$, values as high as $10^7\,{\rm cm}^2/{\rm V\,s}$ are pos-

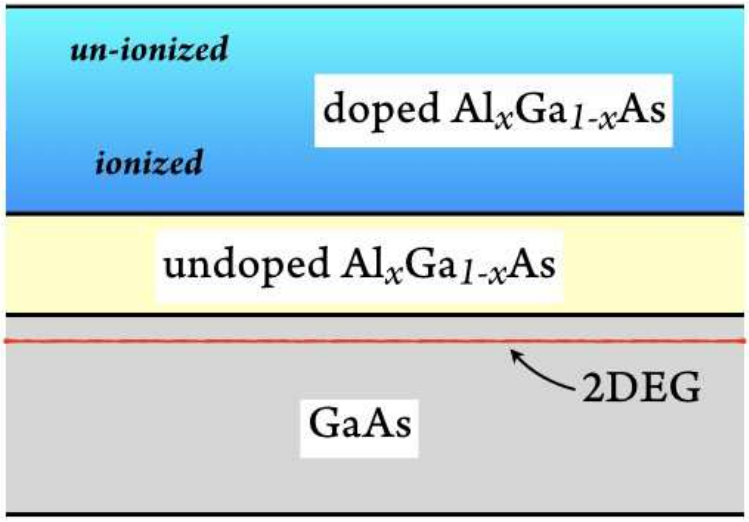


Figure 1.7: GaAs$-$Al$_x$Ga$_{1-x}$As heterojunction.

sible in MBE-fabricated GaAs$-$Al$_x$Ga$_{1-x}$As heterostructures. There are three reasons for this:

(i) MBE (molecular beam epitaxy), as mentioned above, can produce layers which are smooth on an atomic scale. This permits exquisite control of layer thicknesses and doping profiles.

(ii) Use of ternary compounds such as Al$_x$Ga$_{1-x}$As makes for an excellent match in lattice constant across the heterojunction interface, *i.e.* on the order of or better than 1%. By contrast, the Si$-$SiO$_2$ interface is very poor, since SiO$_2$ is a glass.

(iii) By doping the Al$_x$Ga$_{1-x}$As layer far from the interface, Coulomb scattering between inversion layer electrons and dopant ions is suppressed.

Let's consider the chemical potential alignment problem in the case of an $n$ - $n$ heterojunction,

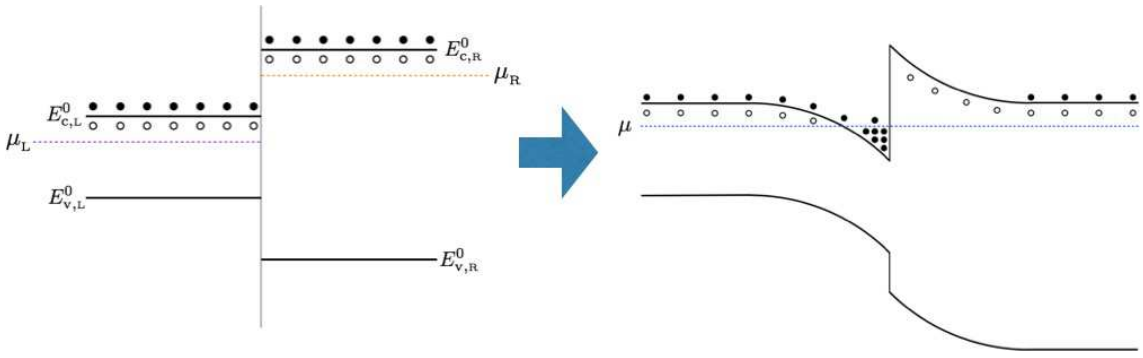


Figure 1.8: Accumulation layer formation in an $n$-$n$ heterojunction.

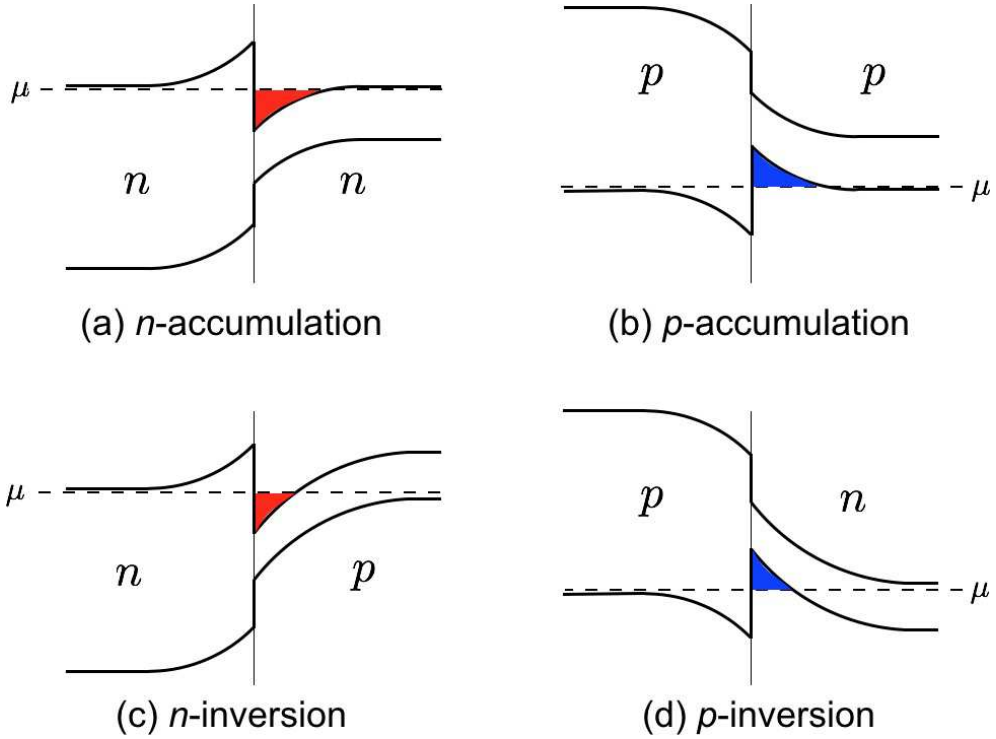


Figure 1.9: Accumulation and inversion in semiconductor heterojunctions. Red regions represent presence of conduction electrons. Blue regions represent presence of valence holes.

sketched in Fig. 1.8. In the GaAs$-$Al$_x$Ga$_{1-x}$As heterojunction, GaAs has the smaller of the two band gaps. Initially there is a mismatch, as depicted in the left panel of the figure. By forming a depletion layer on the side with the larger band gap (Al$_x$Ga$_{1-x}$As), and an accumulation layer on the side with the smaller gap (GaAs), an internal potential $\phi(x)$ is established which aligns the chemical potentials.

Fig. 1.9 shows the phenomena of accumulation and inversion in different possible heterojunctions. There are four possibilities: (a) $n$-$n$, (b) $p$-$p$, (c) $n$-$p$ with the $n$-type material having the larger gap, and (d) $n$-$p$ with the $p$-type material having the larger gap.

### 1.2.3 QM of electron motion normal to 2DEG planes

Consider the case of an $n$-accumulation or $n$-inversion layer as depicted in Fig. 1.9. Let the direction perpendicular to the 2DEG be $\hat{\boldsymbol{z}}$, and let the 2DEG lie on the $z > 0$ side of the interface. Assuming that $\hat{\boldsymbol{z}}$ is a principal axis for the effective mass tensor (with eigenvalue $m_z$), and the magnetic field is along $\pm\hat{\boldsymbol{z}}$, the single electron Hamiltonian is separable into degrees of freedom in the $(x, y)$ plane and those in the $\hat{\boldsymbol{z}}$ direction, *i.e.* $H = H_\perp + H_\parallel$. The eigenvalues and eigenfunctions for $H_\perp$, which governs the planar degrees of freedom, were discussed in §1.3.

Here we consider $H_\parallel$, which we model as

$$H_\parallel = -\frac{\hbar^2}{2m_z}\frac{\partial^2}{\partial z^2} + V(z) \quad , \tag{1.30}$$

with

$$V(z) = -e\phi(z) \approx \begin{cases} 2\pi e\sigma\epsilon^{-1}z & \text{if } z \geq 0 \\ \infty & \text{if } z < 0 \quad . \end{cases} \tag{1.31}$$

Here $\sigma$ is the 2D charge density of the space charge layer and $\epsilon$ the dielectric constant for $z > 0$. Thus, we have a triangular potential.

Next, define the length scale

$$\lambda \equiv \left(\frac{\epsilon\hbar^2}{4\pi\sigma e m_z}\right)^{1/3} \quad , \tag{1.32}$$

the energy scale $\varepsilon_\parallel \equiv \hbar^2/2m_z\lambda^2$, and the dimensionless length $s \equiv z/\lambda$. Then

$$H_\parallel = \varepsilon_\parallel\left(-\frac{\partial^2}{\partial s^2} + s\right) \tag{1.33}$$

with wavefunctions subject to the boundary condition $\varphi(0) = 0$. The solutions are Airy functions. Recall the Airy differential equation,

$$\mathsf{Ai}''(z) - z\,\mathsf{Ai}(z) = 0 \quad . \tag{1.34}$$

Thus, the eigenfunctions of $H_\parallel$ are given by $\varphi_n(z) = \mathsf{Ai}(z + \zeta_n)$, where $\mathsf{Ai}(\zeta_n) = 0$. The first few zeros of $\mathsf{Ai}(z)$ are given by

$$\zeta_1 = -2.3381 \quad , \quad \zeta_2 = -4.0879 \quad , \quad \zeta_3 = -5.5206 \quad , \quad \zeta_4 = -6.7867 \quad , \quad \zeta_5 = -7.9441 \quad . \tag{1.35}$$

The energy eigenvalue corresponding to the eigenfunction $\varphi_n(z)$ is $\mathcal{E}_n = -\zeta_n\,\varepsilon_\parallel$.

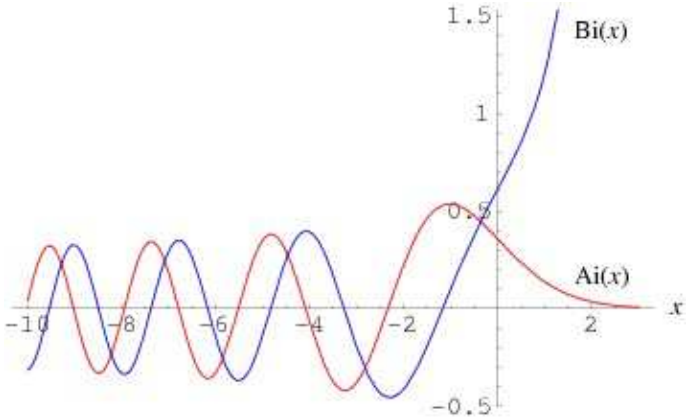


Figure 1.10: Airy functions $\mathsf{Ai}(x)$ and $\mathsf{Bi}(x)$. *Image credit: Wolfram MathWorld.*

## 1.3 Quantization of Planar Motion

### 1.3.1 Cyclotron and guiding-center operators

Initially we shall assume spinless (*i.e.* spin-polarized) electrons. Later on we will include effects of the Zeeman term and explore exchange interactions within a Landau level. The single particle Hamiltonian is then

$$H = \frac{1}{2m}\Big(\boldsymbol{p} + \frac{e}{c}\boldsymbol{A}\Big)^2 + V(\boldsymbol{r}) \quad , \tag{1.36}$$

where $V(\boldsymbol{r})$ is the potential. On a toroidal base space, $V(\boldsymbol{r})$ is a doubly periodic function with spatial periods $\boldsymbol{L}_{1,2}$, and $V(\boldsymbol{r}+\boldsymbol{L}_a) = V(\boldsymbol{r})$ for $a = 1, 2$. We assume $\boldsymbol{B} = -B\hat{\boldsymbol{z}}$ is constant[12]. The cyclotron and guiding-center momenta are defined to be

$$\begin{aligned} \boldsymbol{\pi} &= \boldsymbol{p} + \frac{e}{c}\boldsymbol{A} \\ \boldsymbol{\kappa} &= \boldsymbol{p} + \frac{e}{c}\boldsymbol{A} - \frac{e}{c}\boldsymbol{B}\times\boldsymbol{r} \quad . \end{aligned} \tag{1.37}$$

In component notation, we have $\kappa_\mu = p_\mu + \frac{e}{c}A_\mu - \frac{eB}{c}\,\epsilon_{\mu\rho}\,x_\rho$. The commutators are

$$\begin{aligned} \big[\pi_\mu, \pi_\nu\big] &= \frac{e\hbar}{ic}\left(\partial_\mu A_\nu - \partial_\nu A_\mu\right) = \frac{i\hbar^2}{\ell^2}\,\epsilon_{\mu\nu} \\ \big[\kappa_\mu, \kappa_\nu\big] &= \frac{e\hbar}{ic}\left(\partial_\mu A_\nu - \partial_\nu A_\mu + 2B\,\epsilon_{\mu\nu}\right) = -\frac{i\hbar^2}{\ell^2}\,\epsilon_{\mu\nu} \\ \big[\pi_\mu, \kappa_\nu\big] &= \frac{e\hbar}{ic}\left(\partial_\mu A_\nu - \partial_\nu A_\mu + B\,\epsilon_{\mu\nu}\right) = 0 \quad , \end{aligned} \tag{1.38}$$

where $\ell = \sqrt{\hbar c/eB}$ is the magnetic length. Now we write

$$\boldsymbol{A} = \tfrac{1}{2}By\,\hat{\boldsymbol{x}} - \tfrac{1}{2}Bx\,\hat{\boldsymbol{y}} - \frac{\hbar c}{e}\,\boldsymbol{\nabla}\chi \quad , \tag{1.39}$$

where $\chi(\boldsymbol{r}) = \chi(\boldsymbol{r}+\boldsymbol{L}_a)$ is an arbitrary gauge function which is periodic on the torus[13].

Now define the complexified operators

$$\begin{aligned} \pi = \pi_x + i\pi_y &= \frac{\hbar}{i}\left(\partial_x + i\partial_y\right) + \frac{eB}{2c}\left(y - ix\right) - \hbar\left(\partial_x + i\partial_y\right)\chi \\ &= \frac{2\hbar}{i}\Big(\bar\partial + \frac{z}{4\ell^2} - i\,\bar\partial\chi\Big) = e^{i\chi}\,e^{-z\bar z/4\ell^2}\,\left(-2i\hbar\,\bar\partial\right)e^{z\bar z/4\ell^2}\,e^{-i\chi} \end{aligned} \tag{1.40}$$

[12] By orienting $\boldsymbol{B}$ along $-\hat{\boldsymbol{z}}$, the non-Gaussian part of the lowest Landau level wavefunctions will be holomorphic in $z = x + iy$, rather than in $\bar z = x - iy$.

[13] To demonstrate the manifest gauge covariance of our description, we shall carry around the gauge function $\chi(\boldsymbol{r})$ for a little while. Students should note on their course evaluations that the professor is sensitive to people with diverse gauge preferences.

and

$$\begin{aligned}\kappa = \kappa_x + i\kappa_y &= \frac{\hbar}{i}\,(\partial_x + i\partial_y) - \frac{eB}{2c}\,(y - ix) - \hbar\,(\partial_x + i\partial_y)\,\chi \\ &= \frac{2\hbar}{i}\Big(\bar\partial - \frac{z}{4\ell^2} - i\,\bar\partial\chi\Big) = e^{i\chi}\,e^{z\bar z/4\ell^2}\,(-2i\hbar\,\bar\partial)\,e^{-z\bar z/4\ell^2}\,e^{-i\chi} \quad ,\end{aligned} \tag{1.41}$$

where $\ell = \sqrt{\hbar c/eB}$ is the magnetic length. We have used $z = x + iy$, $\bar z = x - iy$, in which case

$$\partial = \frac{\partial}{\partial z} = \tfrac{1}{2}\big(\partial_x - i\partial_y\big) \qquad , \qquad \bar\partial = \frac{\partial}{\partial \bar z} = \tfrac{1}{2}\big(\partial_x + i\partial_y\big) \quad . \tag{1.42}$$

Note that

$$\partial^\dagger = -\bar\partial \tag{1.43}$$

under Hermitian conjugation. The commutators of the complexified cyclotron and guiding-center operators are given by

$$\big[\pi, \pi^\dagger\big] = \big[\kappa^\dagger, \kappa\big] = \frac{2\hbar^2}{\ell^2} \quad , \tag{1.44}$$

with $\big[\pi, \kappa\big] = \big[\pi^\dagger, \kappa\big] = 0$. We may now define cyclotron and guiding-center ladder operators,

$$\pi = -\frac{i\sqrt{2}\hbar}{\ell}\,a \qquad , \qquad \kappa = \frac{i\sqrt{2}\hbar}{\ell}\,b^\dagger \tag{1.45}$$

with canonical commutators $[a, a^\dagger] = [b, b^\dagger] = 1$. The kinetic term in the Hamiltonian is then

$$H_0 = \frac{\boldsymbol{\pi}^2}{2m} = \frac{\pi^\dagger\pi}{2m} + \frac{\hbar^2}{2m\ell^2} = \hbar\omega_{\rm c}\big(a^\dagger a + \tfrac{1}{2}\big) \quad , \tag{1.46}$$

Note that $H_0$ is cyclic in the guiding-center operators, hence each eigenvalue $\varepsilon_n = (n + \frac{1}{2})\hbar\omega_{\rm c}$ is extensively degenerate. As we shall see below, the degeneracy of each of these *Landau levels* is in the thermodynamic limit equal to $N_{\rm L} = BA/\phi_0$, which is the total magnetic flux through the system in units of the Dirac flux quantum.

We may also define the complexified cyclotron and guiding-center coordinates, $\xi$ and $\mathcal{R}$, as follows:

$$\xi = \frac{i\ell^2}{\hbar}\,\pi = \sqrt{2}\ell\,a \qquad , \qquad \mathcal{R} = -\frac{i\ell^2}{\hbar}\,\kappa = \sqrt{2}\ell\,b^\dagger \quad , \tag{1.47}$$

with $[\mathcal{R}, \mathcal{R}^\dagger] = -2\ell^2$ and $[\xi, \xi^\dagger] = 2\ell^2$. Note then that the complexified position $z = x + iy$ is then given by

$$z = \frac{i\ell^2}{\hbar}\,(\pi - \kappa) = \mathcal{R} + \xi = \sqrt{2}\ell\,(a + b^\dagger) \quad , \tag{1.48}$$

with $\bar z = z^\dagger = \sqrt{2}\ell\,(a^\dagger + b)$. For reference, we also have

$$\partial - i\,\partial\chi = \frac{1}{\sqrt{8}\ell}\,(b - a^\dagger) \qquad , \qquad \bar\partial - i\,\bar\partial\chi = \frac{1}{\sqrt{8}\ell}\,(a - b^\dagger) \quad . \tag{1.49}$$

Finally, the following relations may be useful:

$$a = \sqrt{2}\ell\, e^{i\chi}\, e^{-z\bar{z}/4\ell^2}\, \bar{\partial}\, e^{z\bar{z}/4\ell^2}\, e^{-i\chi} \qquad , \qquad b = \sqrt{2}\ell\, e^{i\chi}\, e^{-z\bar{z}/4\ell^2}\, \partial\, e^{z\bar{z}/4\ell^2}\, e^{-i\chi} \tag{1.50}$$

and

$$a^\dagger = -\sqrt{2}\ell\, e^{i\chi}\, e^{z\bar{z}/4\ell^2}\, \partial\, e^{-z\bar{z}/4\ell^2}\, e^{-i\chi} \qquad , \qquad b^\dagger = -\sqrt{2}\ell\, e^{i\chi}\, e^{z\bar{z}/4\ell^2}\, \bar{\partial}\, e^{-z\bar{z}/4\ell^2}\, e^{-i\chi} \quad . \tag{1.51}$$

*Exercise* : Show that the angular momentum operator satisfies

$$L_z \equiv e^{i\chi}\big(xp_y - yp_x\big)e^{-i\chi} = \hbar\left(b^\dagger b - a^\dagger a\right) \quad . \tag{1.52}$$

### 1.3.2 Landau level projection

Consider the Hamiltonian $H = H_0 + V(\boldsymbol{r})$ confined to the plane. We may write the potential as a Fourier integral

$$V(\boldsymbol{r}) = \int \frac{d^2k}{(2\pi)^2}\, \hat{V}(\boldsymbol{k})\, e^{i\boldsymbol{k}\cdot\boldsymbol{r}} \quad . \tag{1.53}$$

Write $\boldsymbol{r} = \boldsymbol{\mathcal{R}} + \boldsymbol{\xi}$ as a sum over guiding-center and cyclotron coordinates. Since $[\mathcal{R}_\alpha, \xi_\beta] = 0$, we have that

$$\begin{aligned} e^{i\boldsymbol{k}\cdot\boldsymbol{r}} &= e^{i\boldsymbol{k}\cdot\boldsymbol{\mathcal{R}}}\, e^{i\boldsymbol{k}\cdot\boldsymbol{\xi}} = e^{i\boldsymbol{k}\cdot\boldsymbol{\mathcal{R}}}\, e^{-\boldsymbol{k}^2\ell^2/4}\, e^{i\ell k a^\dagger/\sqrt{2}}\, e^{i\ell\bar{k}a/\sqrt{2}} \\ &= e^{-\boldsymbol{k}^2\ell^2/2}\, e^{i\ell\bar{k}b^\dagger/\sqrt{2}}\, e^{i\ell k b/\sqrt{2}}\, e^{i\ell k a^\dagger/\sqrt{2}}\, e^{i\ell\bar{k}a/\sqrt{2}} \quad . \end{aligned} \tag{1.54}$$

We have skipped a few steps. First, we have written $\boldsymbol{k}\cdot\boldsymbol{\xi} = \mathrm{Re}\,(k\xi^\dagger) = \frac{1}{2}(k\xi^\dagger + \bar{k}\xi)$. Then we wrote $\xi = \sqrt{2}\ell\, a$ and $\xi^\dagger = \sqrt{2}\ell\, a^\dagger$ in terms of the cyclotron latter operators. Finally, we wrote $e^{i\ell(ka^\dagger + \bar{k}a)/\sqrt{2}} = e^{-\boldsymbol{k}^2\ell^2/4}\, e^{i\ell k a^\dagger}\, e^{i\ell\bar{k}a/\sqrt{2}}$ using the Baker-Campbell-Hausdorff equality,

$$e^{A+B} = e^A\, e^B\, e^{-\frac{1}{2}[A,B]} \quad , \tag{1.55}$$

which is true when both $A$ and $B$ commute with their commutator $[A, B]$.

Now suppose we *project* the Hamiltonian onto the $n^{\rm th}$ Landau level. This means we evaluate it's expectation value in the cyclotron oscillator state $|\,n\,\rangle$. The result $\langle\, n\,|\,H\,|\,n\,\rangle$ is still an operator, but only in the space of guiding-center states. In other words, it will only involve the operators $b$ and $b^\dagger$ (or $\mathcal{R}$ and $\mathcal{R}^\dagger$). Now we have to roll up our sleeves and do some work. We have

$$\exp\big(i\ell\bar{k}a/\sqrt{2}\big)\,|\,n\,\rangle = \sum_{j=0}^{n} \frac{1}{j!}\left(\frac{i\ell\bar{k}}{\sqrt{2}}\right)^j a^j\,|\,n\,\rangle = \sum_{j=0}^{n} \frac{1}{j!}\left(\frac{i\ell\bar{k}}{\sqrt{2}}\right)^j \sqrt{\frac{n!}{(n-j)!}}\,|\,n-j\,\rangle \tag{1.56}$$

and so

$$\langle\, n\,|\,\exp\big(i\ell k a^\dagger/\sqrt{2}\big)\exp\big(i\ell\bar{k}a/\sqrt{2}\big)\,|\,n\,\rangle = \sum_{j=0}^{n} \frac{1}{j!}\binom{n}{j}\big(-\tfrac{1}{2}\,\boldsymbol{k}^2\ell^2\big)^j \equiv C_n(\boldsymbol{k}\ell) \quad . \tag{1.57}$$

Therefore,

$$V_n(\boldsymbol{\mathcal{R}}) \equiv \langle\, n \,|\, V \,|\, n \,\rangle = \int \frac{d^2k}{(2\pi)^2}\, \hat{V}(\boldsymbol{k})\, e^{i\boldsymbol{k}\cdot\boldsymbol{\mathcal{R}}}\, e^{-\boldsymbol{k}^2\ell^2/4}\, C_n(\boldsymbol{k}\ell) \quad . \tag{1.58}$$

Let's examine what happens for the first few values of $n$:

$$\begin{aligned} C_0(\boldsymbol{k}\ell) &= 1 \\ C_1(\boldsymbol{k}\ell) &= 1 - \tfrac{1}{2}(\boldsymbol{k}\ell)^2 \\ C_2(\boldsymbol{k}\ell) &= 1 - (\boldsymbol{k}\ell)^2 + \tfrac{1}{8}\,(\boldsymbol{k}\ell)^4 \quad , \end{aligned} \tag{1.59}$$

where by $(\boldsymbol{k}\ell)^4$ we mean $|\boldsymbol{k}|^4\ell^4$. Multiplying by the $\exp(-\frac{1}{4}\boldsymbol{k}^2\ell^2)$ factor, and expanding in powers of $\boldsymbol{k}$, we see that the projected potential $V_n(\boldsymbol{\mathcal{R}})$ is given by $V(\boldsymbol{\mathcal{R}})$ plus a series of corrections which can be expressed in terms of powers of the Laplacian $\nabla^2$ acting on $V(\boldsymbol{\mathcal{R}})$. For example,

$$V_0(\boldsymbol{\mathcal{R}}) = V(\boldsymbol{\mathcal{R}}) + \tfrac{1}{4}\,\ell^2\,\nabla^2 V(\boldsymbol{\mathcal{R}}) + \ldots \quad , \tag{1.60}$$

which is the quantum analog of Eqn. 1.28.

Some words of caution are appropriate here. Since $[\mathcal{R}_\alpha, \mathcal{R}_\beta] = -i\ell^2\epsilon_{\alpha\beta}$, we must not be cavalier regarding operator order. To be safe, we might choose to express $V_n(\boldsymbol{\mathcal{R}})$ in some canonical form, such as the *normal ordered* form, in which all the $\mathcal{R}^\dagger$ operators appear to the right of all $\mathcal{R}$ operators[14]. That is, we write

$$\begin{aligned} \exp(i\boldsymbol{k}\cdot\boldsymbol{\mathcal{R}}) &= \exp(\tfrac{i}{2}\bar{k}\mathcal{R})\, \exp(\tfrac{i}{2}k\mathcal{R}^\dagger)\, \exp\big(\tfrac{1}{8}\boldsymbol{k}^2[\mathcal{R},\mathcal{R}^\dagger]\big) \\ &= \exp(-\tfrac{1}{4}\boldsymbol{k}^2\ell^2)\, \exp(\tfrac{i}{2}\bar{k}\mathcal{R})\, \exp(\tfrac{i}{2}k\mathcal{R}^\dagger) \end{aligned} \tag{1.61}$$

in the integrand of Eqn. 1.58. Also, it goes without saying that $\langle\, n \,|\, H_0 \,|\, n \,\rangle = (n + \frac{1}{2})\hbar\omega_{\rm c}$. But I suppose I said it anyway.

### 1.3.3 Landau level mixing

It is apparent that the Hamiltonian $H = H_0 + V(\boldsymbol{r})$ may be written as

$$\begin{aligned} H &= \sum_{n,n=0}^{\infty} H_{nn'}(\boldsymbol{\mathcal{R}})\,|\, n \,\rangle\langle\, n' \,| \\ H_{nn'}(\boldsymbol{\mathcal{R}}) &= (n + \tfrac{1}{2})\hbar\omega_{\rm c}\,\delta_{nn'} + V_{nn'}(\boldsymbol{\mathcal{R}}) \quad , \end{aligned} \tag{1.62}$$

where, for each $\{n, n'\}$, $V_{nn'}(\boldsymbol{\mathcal{R}}) \equiv \langle\, n \,|\, V \,|\, n' \,\rangle$ is an operator in the space of guiding-center degrees of freedom, given by

$$V_{nn'}(\boldsymbol{\mathcal{R}}) = \int \frac{d^2k}{(2\pi)^2}\, \hat{V}(\boldsymbol{k})\, e^{i\boldsymbol{k}\cdot\boldsymbol{\mathcal{R}}}\, e^{-\boldsymbol{k}^2\ell^2/4}\, \langle\, n \,|\, \exp\big(i\ell k a^\dagger/\sqrt{2}\big)\, \exp\big(i\ell\bar{k}a/\sqrt{2}\big)\,|\, n' \,\rangle \quad . \tag{1.63}$$

[14] And, hence, all guiding-center annihilation operators $b$ appear to the right of all guiding-center creation operators $b^\dagger$.

The off-diagonal terms describe *Landau level mixing* processes. For example, if we retain only the $n = 0$ and $n = 1$ LLs, we have the truncated Hamiltonian

$$H_{\rm trunc} = \begin{pmatrix} \frac{1}{2}\hbar\omega_{\rm c} & 0 \\ 0 & \frac{3}{2}\hbar\omega_{\rm c} \end{pmatrix} + \int \frac{d^2k}{(2\pi)^2}\,\hat{V}(\boldsymbol{k})\,e^{i\boldsymbol{k}\cdot\boldsymbol{\mathcal{R}}}\,e^{-\boldsymbol{k}^2\ell^2/4}\begin{pmatrix} 1 & -i\ell\bar{k}/\sqrt{2} \\ i\ell k/\sqrt{2} & 1-\frac{1}{2}\boldsymbol{k}^2\ell^2 \end{pmatrix} \quad . \tag{1.64}$$

Since there is a gap of $\hbar\omega_{\rm c}$ between consecutive Landau levels, LL mixing is usually treated perturbatively.

### 1.3.4 The lowest Landau level

The eigenvalue of $a^\dagger a$ is an integer which corresponds to the Landau level index. For states in the lowest Landau level, we have

$$a\,\psi(\boldsymbol{r}) = 0 \quad \Longrightarrow \quad \psi(\boldsymbol{r}) = e^{i\chi(\boldsymbol{r})}\,e^{-\boldsymbol{r}^2/4\ell^2}\,f(z) \quad , \tag{1.65}$$

where $z = x + iy$. At this point, $f(z)$ is any analytic function. As we shall soon see, periodicity on the torus further constrains the form of $f(z)$.

In zero magnetic field, the density of states (per unit area, per unit energy) is constant:

$$g(\varepsilon, B=0)\,d\varepsilon = \frac{d^2k}{(2\pi)^2} = \frac{k\,dk}{2\pi} \quad \Rightarrow \quad g(\varepsilon) = \frac{m}{2\pi\hbar^2} \tag{1.66}$$

since $\varepsilon = \hbar^2k^2/2m$. When $B$ is finite, the spectrum collapses into discrete Landau levels with energies $\varepsilon_n = (n+\frac{1}{2})\hbar\omega_{\rm c}$. The density of states is

$$g(\varepsilon, B) = \frac{B}{\phi_0}\sum_{n=0}^{\infty}\delta\big(\varepsilon - (n+\tfrac{1}{2})\hbar\omega_{\rm c}\big) \quad . \tag{1.67}$$

The number of Landau levels below energy $E$ is $E/\hbar\omega_{\rm c}$, rounded to the nearest integer. To check the coefficient $B/\phi_0$ in the above expression note that the total number of states per unit area below energy $E$ is then

$$\frac{B}{\phi_0}\cdot\frac{E}{\hbar\omega_{\rm c}} = \frac{mE}{2\pi\hbar^2} \quad , \tag{1.68}$$

which agrees with the $B = 0$ result. Below, we shall count the number of states precisely using a toroidal geometry.

We define the wavefunction $\psi_0(\boldsymbol{r})$ to satisfy $a\psi_0 = b\psi_0 = 0$. Imposing normalization,

$$\psi_0(\boldsymbol{r}) = (2\pi\ell^2)^{-1/2}\,e^{i\chi(\boldsymbol{r})}\,e^{-z\bar{z}/4\ell^2} \quad . \tag{1.69}$$

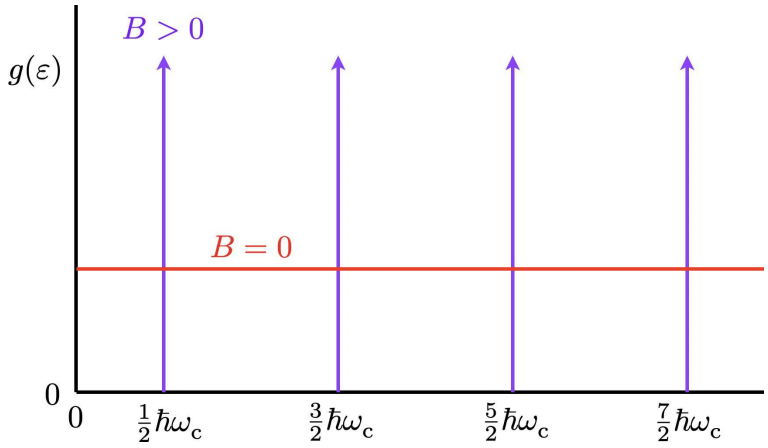


Figure 1.11: Density of states in $d = 2$ for $B = 0$ (red) and $B > 0$ (purple).

A complete and orthonormal set of wavefunctions is given by the collection

$$\begin{aligned}\psi_{m,n}(\boldsymbol{r}) &= \frac{(a^\dagger)^n}{\sqrt{n!}}\,\frac{(b^\dagger)^m}{\sqrt{m!}}\,\psi_0(\boldsymbol{r}) = \frac{\left(-\sqrt{2}\,\ell\right)^{m+n}}{\sqrt{2\pi\ell^2 m!n!}}\,e^{i\chi(\boldsymbol{r})}\,e^{z\bar{z}/4\ell^2}\,\partial^n\,\bar{\partial}^m\,e^{-z\bar{z}/2\ell^2} \\ &= (-1)^n\sqrt{\frac{n!}{2\pi\ell^2 m!}}\left(\frac{z}{\sqrt{2}\ell}\right)^{m-n} L_n^{(m-n)}(z\bar{z}/2\ell^2)\,e^{-z\bar{z}/4\ell^2}e^{i\chi(\boldsymbol{r})} \quad ,\end{aligned} \tag{1.70}$$

where $L_n^{(\alpha)}(x) = \frac{1}{n!}\,x^{-\alpha}\,e^x\,\frac{d^n}{dx^n}\left(x^{n+\alpha}\,e^{-x}\right)$ is an associated Laguerre polynomial. Note that, per Eqn. 1.52, this is also an eigenbasis of angular momentum, *viz.*

$$L_z\,|\,m,n\,\rangle = \hbar(b^\dagger b - a^\dagger a)\,|\,m,n\,\rangle = \hbar(m-n)\,|\,m-n\,\rangle \quad . \tag{1.71}$$

Completeness entails the relation $\sum_{m=0}^\infty\sum_{n=0}^\infty |\,m,n\,\rangle\langle\,m,n\,| = 1$, which is to say

$$\sum_{m=0}^\infty\sum_{n=0}^\infty \psi_{m,n}(\boldsymbol{r})\,\psi^*_{m,n}(\boldsymbol{r}') = \delta(\boldsymbol{r}-\boldsymbol{r}') \quad . \tag{1.72}$$

Note, however, that if we sum over only the states in the lowest Landau level, then

$$\begin{aligned}\sum_{m=0}^\infty \psi_{m,0}(\boldsymbol{r})\,\psi^*_{m,0}(\boldsymbol{r}') &= \frac{e^{i\chi(\boldsymbol{r})}e^{-i\chi(\boldsymbol{r}')}}{2\pi\ell^2}\sum_{m=0}^\infty\frac{1}{m!}\left(\frac{z\bar{z}'}{2\ell^2}\right)^m e^{-z\bar{z}/4\ell^2}e^{-z'\bar{z}'/4\ell^2} \\ &= \frac{e^{i\chi(\boldsymbol{r})}e^{-i\chi(\boldsymbol{r}')}}{2\pi\ell^2}\exp\left(-\frac{z\bar{z}}{4\ell^2}-\frac{z'\bar{z}'}{4\ell^2}+\frac{z\bar{z}'}{2\ell^2}\right) \\ &= \frac{e^{i\chi(\boldsymbol{r})}e^{-i\chi(\boldsymbol{r}')}}{2\pi\ell^2}\exp\left(-\frac{|z-z'|^2}{4\ell^2}+i\,\frac{\mathsf{Im}\,z\bar{z}'}{2\ell^2}\right) \\ &= \frac{e^{i\chi(\boldsymbol{r})}e^{-i\chi(\boldsymbol{r}')}}{2\pi\ell^2}\exp\left(-\frac{(\boldsymbol{r}-\boldsymbol{r}')^2}{4\ell^2}-i\,\frac{\boldsymbol{r}\times\boldsymbol{r}'\cdot\hat{\boldsymbol{z}}}{2\ell^2}\right)\end{aligned} \tag{1.73}$$

rather than $\delta(\boldsymbol{r}-\boldsymbol{r}')$. This tells us that the shortest distance scale on which we can localize an electron in the lowest Landau level is the magnetic length $\ell$.

In the lowest Landau level (LLL), we may write

$$\psi_m(\boldsymbol{r}) = \frac{e^{i\chi(\boldsymbol{r})}}{\sqrt{2\pi\ell^2\, m!}} \left(\frac{z}{\sqrt{2}\ell}\right)^m e^{-z\bar{z}/4\ell^2} \quad . \tag{1.74}$$

This is generally known as the *angular momentum basis*.

### 1.3.5 Landau strip basis

Had we instead chosen the gauge $\boldsymbol{A} = -Bx\hat{\boldsymbol{y}}$ (again corresponding to $\boldsymbol{B} = -B\hat{\boldsymbol{z}}$), then

$$H_0 = \frac{p_x^2}{2m} + \frac{(p_y - \frac{eB}{c}\, x)^2}{2m} \quad . \tag{1.75}$$

There is now translational invariance along $\hat{\boldsymbol{y}}$, hence the wavefunctions may be written as $\psi(x,y) = e^{ik_y y}\,\phi(x)$, with $\phi(x)$ an eigenfunction of

$$H_0(k_y) = \frac{p_x^2}{2m} + \frac{(\hbar k_y - \frac{eB}{c}\, x)^2}{2m} = \frac{p_x^2}{2m} + \tfrac{1}{2}m\omega_{\rm c}^2\,(x - \ell^2 k_y)^2 \quad . \tag{1.76}$$

This is the one-dimensional harmonic oscillator, with eigenfunctions $\phi_n(x-\ell^2 k_y)$, where

$$\phi_n(x) = \frac{1}{\sqrt{2^n n!}}\,(\pi\ell^2)^{-1/4}\,\mathsf{H}_n(x/\ell)\,e^{-x^2/2\ell^2} \quad , \tag{1.77}$$

where $\mathsf{H}_n(u)$ is the $n^{\rm th}$ Hermite polynomial, and corresponding eigenvalues $\varepsilon_n = (n+\frac{1}{2})\hbar\omega_{\rm c}$. The full basis set of wavefunctions as a function of $(x,y)$ is labeled by a discrete Landau level index $n$ and a continuous index $k_y$, *viz.*

$$\psi_{n,k_y}(x,y) = L_y^{-1/2}\,e^{ik_y y}\,\phi_n(x-\ell^2 k_y) \quad . \tag{1.78}$$

On a cylinder $\mathbb{R}\times\mathbb{S}^1$ where $y\in[0,L_y]$, periodic boundary conditions requires $k_y$ to be quantized due to the relation $e^{ik_y L_y}=1$.

### 1.3.6 Magnetic translation operators

The magnetic translation operators (MTOs) are defined as

$$t(\boldsymbol{d}) = \exp(i\boldsymbol{\kappa}\cdot\boldsymbol{d}/\hbar) = \exp\big[(db-\bar{d}b^\dagger)/\sqrt{2}\ell\big] \quad . \tag{1.79}$$

For each $\boldsymbol{d}$, $t^{-1}(\boldsymbol{d}) = t^\dagger(\boldsymbol{d})$, *i.e.* each MTO is a unitary operator. Note also that $\big[t(\boldsymbol{d}), \boldsymbol{\pi}\big] = 0$, so the magnetic translations commute with the kinetic energy, $H_0 = \frac{\pi^2}{2m}$. Acting on any function of the coordinates, we have

$$t(\boldsymbol{d})\,\psi(\boldsymbol{r})\,t^\dagger(\boldsymbol{d}) = \psi(\boldsymbol{r}+\boldsymbol{d}) \quad , \tag{1.80}$$

which is why $t(\boldsymbol{d})$ is a translation operator. It is a worthwhile exercise for the student to show that while $t_0(\boldsymbol{d})\,\psi(\boldsymbol{r}) = \psi(\boldsymbol{r}+\boldsymbol{d})$, where $t_0(\boldsymbol{d}) = \exp(i\boldsymbol{p}\cdot\boldsymbol{d}/\hbar) = \exp(\boldsymbol{d}\cdot\boldsymbol{\nabla})$ is the translation operator with $\boldsymbol{B} = 0$, for the MTOs we have

$$t(\boldsymbol{d})\,\psi(\boldsymbol{r}) = e^{i\chi(\boldsymbol{r})}\,e^{-i\chi(\boldsymbol{r}+\boldsymbol{d})}\,e^{-i\boldsymbol{d}\times\boldsymbol{r}\cdot\hat{\boldsymbol{z}}/2\ell^2}\,\psi(\boldsymbol{r}+\boldsymbol{d}) \quad . \tag{1.81}$$

Due to the magnetic field, two arbitrary magnetic translations do not necessarily commute. Rather,

$$t(\boldsymbol{d}_1)\,t(\boldsymbol{d}_2) = e^{i\hat{\boldsymbol{z}}\cdot\boldsymbol{d}_1\times\boldsymbol{d}_2/2\ell^2}\,t(\boldsymbol{d}_1+\boldsymbol{d}_2) = e^{i\hat{\boldsymbol{z}}\cdot\boldsymbol{d}_1\times\boldsymbol{d}_2/\ell^2}\,t(\boldsymbol{d}_2)\,t(\boldsymbol{d}_1) \quad . \tag{1.82}$$

Thus, $\big[t(\boldsymbol{d}_1), t(\boldsymbol{d}_2)\big] = 0$ if and only if $\hat{\boldsymbol{z}}\cdot\boldsymbol{d}_1\times\boldsymbol{d}_2 = 2\pi\ell^2 q$, where $q$ is an integer.

### 1.3.7 Coherent state wavefunctions

Having tired of carrying the stupid gauge function $\chi(\boldsymbol{r})$ with us for so long, we will now drop it[15], which means working in the symmetric gauge, with $\chi(\boldsymbol{r}) = 0$. Consider again the MTO,

$$\begin{aligned} t(\boldsymbol{R}) &= \exp(i\boldsymbol{\kappa}\cdot\boldsymbol{R}/\hbar) = e^{(Rb-\bar{R}b^\dagger)/\sqrt{2}\ell} \\ &= e^{-R\bar{R}/4\ell^2}e^{-\bar{R}b^\dagger/\sqrt{2}\ell}\,e^{-Rb/\sqrt{2}\ell} \quad , \end{aligned} \tag{1.83}$$

where we have again invoked BCH. The LLL coherent state wavefunction is defined to be[16]

$$|\,\boldsymbol{R}\,\rangle = t^\dagger(\boldsymbol{R})|\,0\,\rangle = e^{-R\bar{R}/4\ell^2}e^{\bar{R}b^\dagger/\sqrt{2}\ell}\,|\,0,0\,\rangle \tag{1.84}$$

It is left as an exercise to the reader to verify the following formulae:

$$\begin{aligned} \varphi_{\boldsymbol{R}}(\boldsymbol{r}) = \langle\,\boldsymbol{r}\,|\,t^\dagger(\boldsymbol{R})\,|\,0\,\rangle &= \frac{1}{\sqrt{2\pi\ell^2}}\,e^{-i\boldsymbol{r}\times\boldsymbol{R}\cdot\hat{\boldsymbol{z}}/2\ell^2}e^{-(\boldsymbol{r}-\boldsymbol{R})^2/4\ell^2} \\ \langle\,\boldsymbol{R}\,|\,\boldsymbol{R}'\,\rangle &= \exp\left(-\frac{(\boldsymbol{R}-\boldsymbol{R}')^2}{4\ell^2} - i\,\frac{\boldsymbol{R}\times\boldsymbol{R}'\cdot\hat{\boldsymbol{z}}}{2\ell^2}\right) \\ \int\frac{d^2R}{2\pi\ell^2}\,|\,\boldsymbol{R}\,\rangle\langle\,\boldsymbol{R}\,| &= \sum_{m=0}^{\infty}|\,m,0\,\rangle\langle\,m,0\,| \equiv \Pi_0 \quad . \end{aligned} \tag{1.85}$$

[15] As we are feeling persons, we hope that the gauge function manages to find its way back home to its loved ones, whom it can regale with heroic stories of having served alongside other important factors in various operators and wavefunctions.

[16] When confining our attention to the LLL, it is convenient to drop the cyclotron quantum number and write the guiding center vacuum state simply as $|\,0\,\rangle$ rather than $|\,0,0\,\rangle$.

Thus, the coherent state wavefunction $\varphi_{\boldsymbol{R}}(\boldsymbol{r})$ is Gaussianly localized about $\boldsymbol{r} = \boldsymbol{R}$, but contains phase information as well. The coherent states admit a resolution of unity within the lowest Landau level, or indeed for any Landau level if we define $|\,\boldsymbol{R}, n\,\rangle \equiv t^\dagger(\boldsymbol{R})\,|\,0, n\,\rangle$, in which case

$$\int \frac{d^2R}{2\pi\ell^2}\,|\,\boldsymbol{R}, n\,\rangle\langle\,\boldsymbol{R}, n\,| = \sum_{m=0}^{\infty} |\,m, n\,\rangle\langle\,m, n\,| \equiv \Pi_n \tag{1.86}$$

is the projector onto the $n^{\rm th}$ Landau level.

## 1.4 Landau Levels in Graphene

### 1.4.1 Quick overview

First we will skip all the details and quickly derive the spectrum of the Landau levels in graphene, which is different than for the case in ballistic 2DEGs. Then we will circle back and build a model of what graphene actually *is* at the atomic level and validate everything we find. We'll even allow for lattice *strain* and see how it can generate a *pseudomagnetic field* which in experiments can be on the order of 300 Tesla!

The two-dimensional Dirac Hamiltonian is

$$H = \hbar v_{\rm F}\,\boldsymbol{\sigma}\cdot\boldsymbol{q} = \hbar v_{\rm F}\begin{pmatrix} 0 & q_x - iq_y \\ q_x + iq_y & 0 \end{pmatrix} \quad , \tag{1.87}$$

where $v_{\rm F} \approx c/300$ is the Fermi velocity in graphene. At this point, the electron spin is a spectator, and the Pauli matrix structure here is in the *isospin* space associated with the two triangular sublattices of the hexagonal graphene structure. To describe the Zeeman term, or, more interestingly, spin-orbit effects, we would need to invoke a second set of Pauli matrices $(\tau^x, \tau^y, \tau^z)$ acting on the spin degrees of freedom[17]. In the presence of a magnetic field, the minimal coupling prescription applies[18], and we have

$$H(\boldsymbol{A}) = v_{\rm F}\begin{pmatrix} 0 & \pi^\dagger \\ \pi & 0 \end{pmatrix} \tag{1.88}$$

where $\pi = \hbar(q_x + iq_y) + \frac{e}{c}(A_x + iA_y)$ as before. We then have

$$[\pi, \pi^\dagger] = \frac{2\hbar e}{c}\big(\partial_y A_x - \partial_x A_y\big) = -\frac{\hbar e B_z}{c} \quad , \tag{1.89}$$

[17] In graphene, spin-orbit effects are very weak due to the low $Z$ value of carbon. In higher $Z$ systems, spin-orbit effects are crucial and may give rise to topological insulator behavior.

[18] The validity of the Peierls substitution $E_n(\boldsymbol{k}) \to E_n(\boldsymbol{k} + \frac{e}{\hbar c}\boldsymbol{A})$ for Bloch electrons is a nontrivial matter and was first established by W. Kohn, *Phys. Rev.* **115**, 1460 (1959).

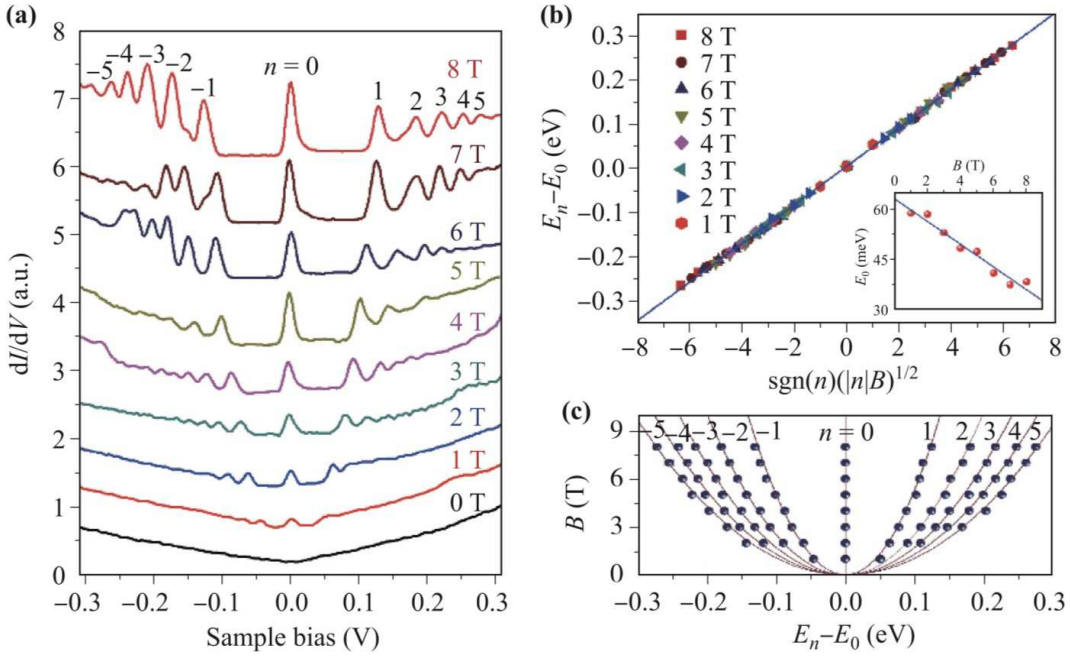


Figure 1.12: Landau levels in monolayer graphene, as reported by L.-J. Yin *et al.*, *Front. Phys.* **12**, 127208 (2017). Left: Differential conductance spectra of a graphene monolayer on a graphite surface as a function of bias voltage and for several values of magnetic field. Right: The LL dispersion is in excellent agreement with the 2D Dirac theory.

and assuming $\boldsymbol{B} = -B\hat{\boldsymbol{z}}$, we again recover $[\pi, \pi^\dagger] = 2\hbar^2\ell^{-2}$ with $\ell = \sqrt{\hbar c/eB}$ the magnetic length. Defining the cyclotron ladder operators $a = \ell\pi/\sqrt{2}\,\hbar$ and $a^\dagger = \ell\pi^\dagger/\sqrt{2}\,\hbar$ as before[19], we have

$$H = \frac{\sqrt{2}\hbar v_{\rm F}}{\ell}\begin{pmatrix} 0 & a^\dagger \\ a & 0 \end{pmatrix} \quad . \tag{1.90}$$

Note that the guiding center operators $b$ and $b^\dagger$ are cyclic in $H$, hence there is an extensive degeneracy of each Landau level.

It is easy to see that the eigenvectors of $H$, expressed in terms of the cyclotron oscillator states $|\,n\,\rangle$, are given by

$$|\,\Psi_0\,\rangle = \begin{pmatrix} |\,0\,\rangle \\ 0 \end{pmatrix} \qquad , \qquad |\,\Psi_{\pm,n}\,\rangle = \frac{1}{\sqrt{2}}\begin{pmatrix} |\,n\,\rangle \\ \pm|\,n-1\,\rangle \end{pmatrix} \tag{1.91}$$

where $n \in \mathbb{Z}_+ = \{1, 2, 3, \ldots\}$. The corresponding eigenvalues are

$$E_0 = 0 \qquad , \qquad E_{\pm,n} = \pm\sqrt{2n}\,\frac{\hbar v_{\rm F}}{\ell} \quad . \tag{1.92}$$

[19]Not quite as before. In Eqn. 1.45 there is an extra factor of $i$ which we find convenient to remove here.

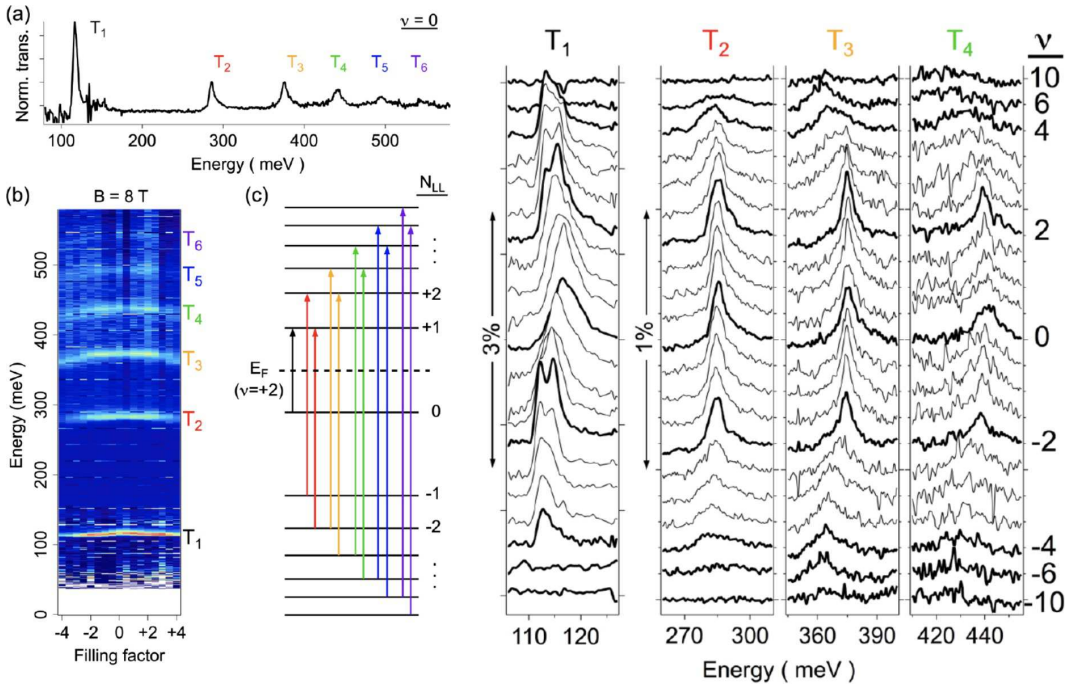


Figure 1.13: Infrared cyclotron resonance data on high mobility monolayer graphene by B. J. Russell *et al.*, *Phys. Rev. Lett.* **120**, 047401 (2018). The interband LL transitions are measured at fixed $B = 8T$. The density $n$ and thus the filling fraction $\nu$ are varied by an applied gate voltage. Left: (a) Normalized transition at $\nu = 0$, showing the first six resonances. (b) Color map of resonances *vs.* filling fraction, showing small variations even at fixed $\omega_c$ . (c) Schematic of inter-LL transitions. Right: Dependence of the first four inter-LL transitions on filling fraction. Note each filled LL index corresponds to $\Delta\nu = 2$ because of spin degeneracy.

We now note several important distinctions between this spectrum and that for the ballistic case derived in §1.3:

(i) The LL spectrum is particle-hole symmetric, with a zero energy LL and symmetrically arranged levels at $E_{-,n} = -E_{+,n}$.

(ii) Rather than the ballistic LL spectrum $E_n = (n + \frac{1}{2})\hbar\omega_c$ where $\omega_c = eB/mc$, the Dirac LL spectrum is not evenly spaced nor is it linear in $B$. Instead, energies grow with LL index and field as $\sqrt{|nB|}$.

It turns out that this is only half the story, because there is another Dirac point in the graphene band structure where the corresponding long wavelength Hamiltonian is $H' = \sigma^x H \sigma^x$. The eigenspectra of these two 2D Dirac Hamiltonians are therefore identical.

Scanning probe measurements in monolayer graphene shown in Fig. **??** reveal an excellent fit to the Dirac dispersion of Eqn. 1.92. More careful investigations, however, from infrared

cyclotron resonance (Fig. 1.13) show small variations in inter-LL resonance energies at fixed $B$ as a function of the filling fraction $\nu$. These are attributed to interaction effects, which we shall study later on in these notes.

OK, now let's derive all this stuff from scratch.

### 1.4.2 Direct and reciprocal lattice

Graphene is a two-dimensional form of pure carbon arranged in a honeycomb lattice, where each site is threefold coordinated. The electronic configuration of C is $[\mathrm{He}]\,2\mathrm{s}^2\,2\mathrm{p}^2$. The $2\mathrm{s}$ and $2\mathrm{p}$ orbitals engage in $\mathrm{sp}^2$ hybridization, forming on each carbon atom three planar orbitals oriented $120^\circ$ from each other. These engage in covalent bonding with each of a given C atom's three neighbors. The remaining fourth electron is in a $p_z$ state – the so-called $\pi$ orbital. The simplest model of graphene considers as inert the [He] core and the covalently bonded $\mathrm{sp}^2$ orbitals and focuses on the remaining single $\pi$ electron per site.

The honeycomb lattice is a triangular Bravais lattice with a two element basis. The Bravais lattice sites are located at $\boldsymbol{R} = n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2$, with elementary direct lattice vectors

$$\boldsymbol{a}_1 = a\left(\tfrac{1}{2}\hat{\boldsymbol{x}} - \tfrac{\sqrt{3}}{2}\hat{\boldsymbol{y}}\right) \quad , \quad \boldsymbol{a}_2 = a\left(\tfrac{1}{2}\hat{\boldsymbol{x}} + \tfrac{\sqrt{3}}{2}\hat{\boldsymbol{y}}\right) \quad , \tag{1.93}$$

as shown in the left panel of Fig. 1.14. The basis vectors are then $0$ and $\boldsymbol{\delta}_1 = \delta\hat{\boldsymbol{y}}$, where $\delta = a/\sqrt{3}$ is the spacing between C atoms. It is useful to define the two other nearest neighbor vectors $\boldsymbol{\delta}_{2,3}$ as shown in the figure, in which case

$$\boldsymbol{\delta}_1 = \tfrac{1}{3}(-\boldsymbol{a}_1 + \boldsymbol{a}_2) \quad , \quad \boldsymbol{\delta}_2 = \tfrac{1}{3}(2\boldsymbol{a}_1 + \boldsymbol{a}_2) \quad , \quad \boldsymbol{\delta}_3 = \tfrac{1}{3}(-\boldsymbol{a}_1 - 2\boldsymbol{a}_2) \quad . \tag{1.94}$$

Note that $\boldsymbol{\delta}_1 + \boldsymbol{\delta}_2 + \boldsymbol{\delta}_3 = 0$. The reciprocal lattice is triangular, with elementary reciprocal lattice vectors

$$\boldsymbol{b}_1 = \frac{4\pi}{a\sqrt{3}}\left(\tfrac{\sqrt{3}}{2}\hat{\boldsymbol{x}} - \tfrac{1}{2}\hat{\boldsymbol{y}}\right) \quad , \quad \boldsymbol{b}_2 = \frac{4\pi}{a\sqrt{3}}\left(\tfrac{\sqrt{3}}{2}\hat{\boldsymbol{x}} + \tfrac{1}{2}\hat{\boldsymbol{y}}\right) \quad . \tag{1.95}$$

Note that $\boldsymbol{a}_i \cdot \boldsymbol{b}_j = 2\pi\delta_{ij}$. Let the tight binding hopping matrix elements along the $\boldsymbol{\delta}_1$, $\boldsymbol{\delta}_2$, and $\boldsymbol{\delta}_3$ directed links be $-t_1$, $-t_2$, and $-t_3$, respectively. Writing the fermionic creation operators for electrons at the A and B sites in unit cell $\boldsymbol{R}$ as $a^\dagger_{\boldsymbol{R}}$ and $b^\dagger_{\boldsymbol{R}}$, the tight binding Hamiltonian is then

$$\begin{aligned} H &= -\sum_{\boldsymbol{R}} \Big\{ \left(t_1\, a^\dagger_{\boldsymbol{R}}\, b_{\boldsymbol{R}} + t_2\, a^\dagger_{\boldsymbol{R}}\, b_{\boldsymbol{R}+\boldsymbol{a}_1} + t_3\, a^\dagger_{\boldsymbol{R}}\, b_{\boldsymbol{R}-\boldsymbol{a}_2}\right) + \mathrm{H.c.} \Big\} \\ &= -t\sum_{\boldsymbol{k}} \Big\{ a^\dagger_{\boldsymbol{k}}\, b_{\boldsymbol{k}} \left(u + v\, e^{i\boldsymbol{k}\cdot\boldsymbol{a}_1} + w\, e^{-i\boldsymbol{k}\cdot\boldsymbol{a}_2}\right) + \mathrm{H.c.} \Big\} \quad , \end{aligned} \tag{1.96}$$

where we have written $t_1 \equiv ut$, $t_2 \equiv vt$, and $t_3 \equiv wt$, with $t$ the fundamental hopping energy scale and $u$, $v$, and $w$ all dimensionless. We have diagonalized $H$ within each crystal momentum $\boldsymbol{k}$ sector via the relations

$$a_{\boldsymbol{R}} = \frac{1}{\sqrt{N_{\mathrm{c}}}}\sum_{\boldsymbol{k}} a_{\boldsymbol{k}}\, e^{i\boldsymbol{k}\cdot\boldsymbol{R}} \quad , \quad a_{\boldsymbol{k}} = \frac{1}{\sqrt{N_{\mathrm{c}}}}\sum_{\boldsymbol{R}} a_{\boldsymbol{R}}\, e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \quad , \tag{1.97}$$

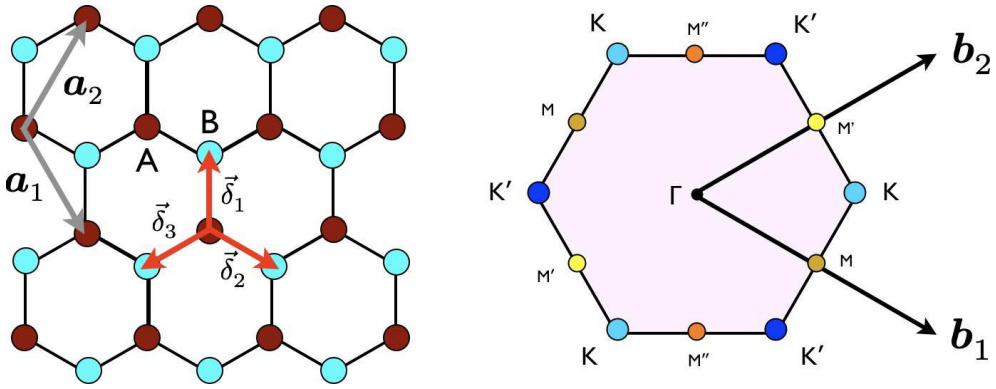


Figure 1.14: The honeycomb lattice is a triangular Bravais lattice with a two element basis (A and B). Left: Real space lattice, with elementary direct lattice vectors $\boldsymbol{a}_{1,2}$. Right: Brillouin zone, with elementary reciprocal lattice vectors $\boldsymbol{b}_{1,2}$ and high symmetry points identified.

where $N_{\rm c}$ is the number of unit cells, and with corresponding definitions on the B sites. Thus $\{a_{\boldsymbol{R}}, a^\dagger_{\boldsymbol{R}'}\} = \delta_{\boldsymbol{R},\boldsymbol{R}'}$ and $\{a_{\boldsymbol{k}}, a^\dagger_{\boldsymbol{k}'}\} = \delta_{\boldsymbol{k},\boldsymbol{k}'}$ . We now define

$$\gamma_{\boldsymbol{k}} \equiv u + v\, e^{i\boldsymbol{k}\cdot\boldsymbol{a}_1} + w\, e^{-i\boldsymbol{k}\cdot\boldsymbol{a}_2} \quad , \tag{1.98}$$

which allows us to write

$$H = -t \sum_{\boldsymbol{k}} \begin{pmatrix} a^\dagger_{\boldsymbol{k}} & b^\dagger_{\boldsymbol{k}} \end{pmatrix} \begin{pmatrix} 0 & \gamma_{\boldsymbol{k}} \\ \gamma^*_{\boldsymbol{k}} & 0 \end{pmatrix} \begin{pmatrix} a_{\boldsymbol{k}} \\ b_{\boldsymbol{k}} \end{pmatrix} \quad . \tag{1.99}$$

The energy eigenvalues are this $E_{\boldsymbol{k},\pm} = \pm t\,|\gamma_{\boldsymbol{k}}|$ . Again, electron spin is a spectator at this point, so each level is doubly degenerate in the absence of, *e.g.*, a Zeeman term[20].

It is convenient to write the crystal wavevector $\boldsymbol{k}$ as

$$\boldsymbol{k} \equiv \frac{\theta_1}{2\pi}\,\boldsymbol{b}_1 + \frac{\theta_2}{2\pi}\,\boldsymbol{b}_2 \tag{1.100}$$

in which case $\exp(i\boldsymbol{k}\cdot\boldsymbol{a}_{1,2}) = \exp(i\theta_{1,2})$ .

### 1.4.3 Long wavelength Hamiltonian

When $u = v = w = 1$ and all the hopping amplitudes are equal to $t$, there are Dirac points at the two inequivalent zone corners $\boldsymbol{K}$ and $\boldsymbol{K}' = -\boldsymbol{K}$, where $\boldsymbol{K} = \frac{1}{3}(\boldsymbol{b}_1 + \boldsymbol{b}_2)$. At $\boldsymbol{k} = \boldsymbol{K}$, one has $\gamma_{\boldsymbol{K}} = 1 + e^{2\pi i/3} + e^{-2\pi i/3} = 0$, and similarly for $\boldsymbol{k} = \boldsymbol{K}'$. What happens when the hopping

[20]For a detailed description of the electronic properties of graphene, see A. H. Castro Neto *et al.*, *Rev. Mod. Phys.* **81**, 109 (2009).

amplitudes are not all the same? Having chosen the overall energy scale $t$, we may fix the sum $u+v+w=3$ and choose to write

$$u = 1 + \tfrac{1}{2}\,\varepsilon_1 + \tfrac{1}{6}\,\varepsilon_2 \qquad , \qquad v = 1 - \tfrac{1}{2}\,\varepsilon_1 + \tfrac{1}{6}\,\varepsilon_2 \qquad , \qquad w = 1 - \tfrac{1}{3}\,\varepsilon_2 \quad . \tag{1.101}$$

We will write $\boldsymbol{k} = \boldsymbol{K} + \boldsymbol{q}$, with the deviation $\boldsymbol{q}$ from the Dirac point presumed to be small. Similarly, we presume that the dimensionless hopping anisotropies $\varepsilon_{1,2}$ are also small. Working to first order in smallness for each, we find

$$\begin{aligned}\gamma_{\boldsymbol{K}+\boldsymbol{q}} &= \big(1 + \tfrac{1}{2}\varepsilon_2 + \tfrac{1}{6}\,\varepsilon_3\big) + \big(1 - \tfrac{1}{2}\varepsilon_2 + \tfrac{1}{6}\,\varepsilon_3\big)\,e^{2\pi i/3}\,\big(1 + i\boldsymbol{q}\cdot\boldsymbol{a}_1 + \ldots\big) \\ &\qquad + \big(1 - \tfrac{1}{3}\,\varepsilon_3\big)\,e^{-2\pi i/3}\,\big(1 - i\boldsymbol{q}\cdot\boldsymbol{a}_2 + \ldots\big) \\ &= -\tfrac{\sqrt{3}}{2}\,(q_x - iq_y)\,a + \tfrac{1}{4}\,(3\varepsilon_1 + \varepsilon_2) - i\,\tfrac{\sqrt{3}}{4}\,(\varepsilon_1 - \varepsilon_2) + \ldots \\ &= -\tfrac{\sqrt{3}}{2}\,a\,\big(q_x - iq_y + Q_x - iQ_y\big) \quad ,\end{aligned} \tag{1.102}$$

where

$$Q_x \equiv -\frac{1}{2\sqrt{3}\,a}\,(3\varepsilon_1 + \varepsilon_2) \qquad , \qquad Q_y = \frac{1}{2a}\,(\varepsilon_2 - \varepsilon_1) \quad . \tag{1.103}$$

We may even allow $\varepsilon_{1,2}$ to vary slowly in space, in which case we must also impose the canonical commutation relations $\big[q_\alpha, x_\beta\big] = -i\,\delta_{\alpha\beta}$. Note that we may now read off $v_{\rm F} = \sqrt{3}\,ta/2\hbar$.

In the presence of an external gauge field, the prescription is the usual minimal coupling, *viz.*

$$\gamma_{\boldsymbol{K}+\boldsymbol{q}} = -\tfrac{\sqrt{3}}{2}\,a\left(q_x - iq_y + Q_x - iQ_y + \frac{e}{\hbar c}\,A_x - i\,\frac{e}{\hbar c}\,A_y\right) \quad , \tag{1.104}$$

Note that $q_x = -i\,\partial_x$ and $q_y = -i\,\partial_y$ are in fact differential operators for our purposes. Further defining

$$\theta = -\gamma_{\boldsymbol{K}+\boldsymbol{q}} = \tfrac{\sqrt{3}}{2}\,a\left[\left(q_x + Q_x + \frac{eA_x}{\hbar c}\right) - i\left(q_y + Q_y + \frac{eA_y}{\hbar c}\right)\right] \quad , \tag{1.105}$$

we derive

$$\big[\theta, \theta^\dagger\big] = \tfrac{3}{2}\,a^2\left(\partial_x Q_y - \partial_y Q_x + \frac{eB_z}{\hbar c}\right) \equiv \pm r^2 \quad . \tag{1.106}$$

We presume that $r^2$ is a constant, independent of space. If the sign on the RHS is positive, define $\theta \equiv r\alpha$, with $[\alpha, \alpha^\dagger] = 1$. Else, if negative, define $\theta \equiv r\alpha^\dagger$. In the former case, we have $\gamma_{\boldsymbol{K}+\boldsymbol{q}} = -r\alpha$ and in the latter $\gamma_{\boldsymbol{K}+\boldsymbol{q}} = -r\alpha^\dagger$. Thus,

$$H_{\boldsymbol{K}} = t\begin{pmatrix} 0 & \theta \\ \theta^\dagger & 0 \end{pmatrix} \qquad , \qquad H_{\boldsymbol{K},+} = rt\begin{pmatrix} 0 & \alpha \\ \alpha^\dagger & 0 \end{pmatrix} \qquad , \qquad H_{\boldsymbol{K},-} = rt\begin{pmatrix} 0 & \alpha^\dagger \\ \alpha & 0 \end{pmatrix} \quad . \tag{1.107}$$

Let's work out the eigenspectrum for $H_+$; the $H_-$ case is equivalent since $H_- = \sigma^x\,H_+\,\sigma^x$. Let $\alpha\,|\,0\,\rangle = 0$ and $|\,n\,\rangle \equiv (\alpha^\dagger)^n\,|\,0\,\rangle/\sqrt{n!}$ with $n \in \mathbb{Z}_+$. Then it is easy to see that

$$H_+\,|\,\Psi_0\,\rangle = 0 \qquad , \qquad H_+\,|\,\Psi_{\pm,n}\,\rangle = \pm\sqrt{n}\,r\,t\,|\,\Psi_{\pm,n}\,\rangle \tag{1.108}$$

where

$$|\,\Psi_0\,\rangle = \begin{pmatrix} |\,0\,\rangle \\ 0 \end{pmatrix} \quad , \quad |\,\Psi_{\pm,n}\,\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} |\,n\,\rangle \\ \pm|\,n-1\,\rangle \end{pmatrix} \quad . \tag{1.109}$$

The corresponding eigenstates of $H_-$ are then given by $\sigma^x\,|\,\Psi_0\,\rangle$ and $\sigma^x\,|\,\Psi_{n,\pm}\,\rangle$. These are the Landau levels of the 2D Dirac Hamiltonian. As with the familiar case with continuum Landau levels from a ballistic dispersion, the Dirac Hamiltonian Landau levels are also massively degenerate. Note that the Landau level energy varies as $\sqrt{n}$ and not linearly in $n$ as in the ballistic case!

### 1.4.4 The $\boldsymbol{K}'$ valley

Consider now the other inequivalent zone corner, located at $\boldsymbol{K}' = -\boldsymbol{K}$. We then have

$$\begin{aligned} \gamma_{-\boldsymbol{K}+\boldsymbol{q}} &= u + v\,e^{-2\pi i/3}\,e^{i\boldsymbol{q}\cdot\boldsymbol{a}_1} + w\,e^{2\pi i/3}\,e^{-i\boldsymbol{q}\cdot\boldsymbol{a}_2} \\ &= \tfrac{\sqrt{3}}{2}\,(q_x + iq_y)\,a + \tfrac{1}{4}(3\varepsilon_1 + \varepsilon_2) + i\,\tfrac{\sqrt{3}}{4}\,(\varepsilon_1 - \varepsilon_2) \\ &= \tfrac{\sqrt{3}}{2}\,a\,\big(q_x + iq_y - Q_x - iQ_y\big) \quad . \end{aligned} \tag{1.110}$$

Again we include the electromagnetic gauge field via minimal coupling,

$$\gamma_{-\boldsymbol{K}+\boldsymbol{q}} = \tfrac{\sqrt{3}}{2}\,a\left(q_x + iq_y - Q_x - iQ_y + \frac{e}{\hbar c}\,A_x + i\,\frac{e}{\hbar c}\,A_y\right) \quad , \tag{1.111}$$

and we define

$$\theta = -\gamma^*_{-\boldsymbol{K}+\boldsymbol{q}} = -\tfrac{\sqrt{3}}{2}\,a\left[\left(q_x - Q_x + \frac{eA_x}{\hbar c}\right) - i\left(q_y - Q_y + \frac{eA_y}{\hbar c}\right)\right] \quad , \tag{1.112}$$

we derive

$$\big[\theta, \theta^\dagger\big] = \tfrac{3}{2}\,a^2\left(-\,\partial_x Q_y + \partial_y Q_x + \frac{eB_z}{\hbar c}\right) \equiv \pm s^2 \quad . \tag{1.113}$$

The Hamiltonian in the $\boldsymbol{K}'$ valley is $H_{\boldsymbol{K}'} = t\begin{pmatrix} 0 & \theta^\dagger \\ \theta & 0 \end{pmatrix}$. Note that the magnetic flux has the same sign in Eqns. 1.106 and 1.113, but that the contribution from the hopping anisotropy is reversed. This is because magnetic flux breaks time reversal symmetry and hopping anisotropy does not, even though the spatially varying hopping anisotropy, which is due to strain, can generate Landau levels. What happens is that the Landau levels in the $\boldsymbol{K}$ valley are the time reverse states of the Landau levels in the $\boldsymbol{K}'$ valley.

### 1.4.5 Strain and pseudomagnetic fields

We now need a model for the distortions $\varepsilon_{1,2}$. Consider the case of *triaxial strain* where the local displacement at position $(x, y)$ is given by $\boldsymbol{u}(x, y)$, where

$$u_x = 2\eta xy \quad , \quad u_y = \eta(x^2 - y^2) \quad . \tag{1.114}$$

Here $\eta$ has dimensions of inverse length. Since hopping amplitudes typically vary exponentially with distance, we write, phenomenologically,

$$t_{\boldsymbol{\delta}} = t\,\exp\left(\frac{\delta_0 - |\boldsymbol{\delta}|}{\lambda}\right) \quad , \tag{1.115}$$

where $\delta_0 = a/\sqrt{3}$ is the unstrained nearest neighbor C-C bond length and $\lambda$ is a constant with dimensions of length corresponding to the transverse extent of the atomic $\pi$ orbital. Under strain, we have

$$\boldsymbol{\delta} \to \boldsymbol{\delta}' = \boldsymbol{\delta} + \boldsymbol{u}(\boldsymbol{R} + \boldsymbol{\delta}) - \boldsymbol{u}(\boldsymbol{R}) \quad , \tag{1.116}$$

and therefore to lowest nontrivial order

$$\begin{aligned} \delta'_x &= \delta_x + 2\eta\,(X\delta_y + Y\delta_x) \\ \delta'_y &= \delta_y + 2\eta\,(X\delta_x - Y\delta_y) \quad , \end{aligned} \tag{1.117}$$

with $\boldsymbol{R} = (X, Y)$. From these relations, we obtain the extensions

$$|\boldsymbol{\delta}'| - \delta_0 = \Delta_x(\boldsymbol{\delta})\,X + \Delta_y(\boldsymbol{\delta})\,Y \tag{1.118}$$

with

$$\Delta_x(\boldsymbol{\delta}) = 2\eta\delta_0 \cdot 2\hat{\delta}_x\hat{\delta}_y \qquad , \qquad \Delta_y(\boldsymbol{\delta}) = 2\eta\delta_0 \cdot \big(\hat{\delta}_x^2 - \hat{\delta}_y^2\big) \quad . \tag{1.119}$$

Thus,

$$\boldsymbol{\Delta}(\boldsymbol{\delta}_1) = -2\eta\delta_0\,\hat{\boldsymbol{y}} \quad , \quad \boldsymbol{\Delta}(\boldsymbol{\delta}_2) = 2\eta\delta_0\big(-\tfrac{\sqrt{3}}{2}\,\hat{\boldsymbol{x}} + \tfrac{1}{2}\,\hat{\boldsymbol{y}}\big) \quad , \quad \boldsymbol{\Delta}(\boldsymbol{\delta}_3) = 2\eta\delta_0\big(\tfrac{\sqrt{3}}{2}\,\hat{\boldsymbol{x}} + \tfrac{1}{2}\,\hat{\boldsymbol{y}}\big) \quad . \tag{1.120}$$

We now have

$$t(\boldsymbol{\delta}) = t\,\exp\left(\frac{\delta_0 - |\boldsymbol{\delta}|}{\lambda}\right) = t\big(1 - \boldsymbol{\Delta}(\boldsymbol{\delta}) \cdot \boldsymbol{R}/\lambda + \ldots\big) \quad , \tag{1.121}$$

which says

$$\begin{aligned} t_1(\boldsymbol{R}) &= \big(1 + \tfrac{1}{2}\,\varepsilon_1 + \tfrac{1}{6}\,\varepsilon_2\big)\,t = \big(1 + 2\eta\,\delta_0\,Y\lambda^{-1} + \ldots\big)\,t \\ t_2(\boldsymbol{R}) &= \big(1 - \tfrac{1}{2}\,\varepsilon_1 + \tfrac{1}{6}\,\varepsilon_2\big)\,t = \big(1 + \sqrt{3}\,\eta\,\delta_0\,X\lambda^{-1} - \eta\,\delta_0\,Y\lambda^{-1} + \ldots\big)\,t \\ t_3(\boldsymbol{R}) &= \big(1 - \tfrac{1}{3}\,\varepsilon_2\big)\,t = \big(1 - \sqrt{3}\,\eta\,\delta_0\,X\lambda^{-1} - \eta\,\delta_0\,Y\lambda^{-1} + \ldots\big)\,t \quad . \end{aligned} \tag{1.122}$$

Solving for $\varepsilon_{1,2}$, we obtain

$$\begin{aligned} \varepsilon_1 &= \sqrt{3}\,\eta\delta_0\,\big(-X + \sqrt{3}\,Y\big)\lambda^{-1} \\ \varepsilon_2 &= 3\eta\delta_0\,\big(\sqrt{3}\,X + Y\big)\lambda^{-1} \end{aligned} \tag{1.123}$$

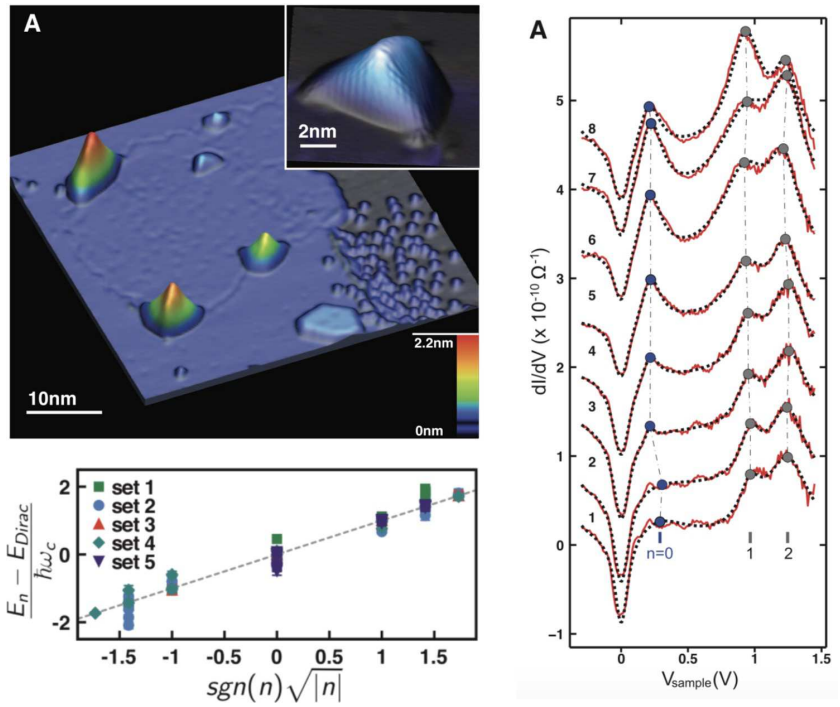


Figure 1.15: Strain-induced pseudo-Landau levels in graphene, reported by N. Levy *et al.*, *Science* **329**, 544 (2010). Upper left: STM images taken at $7.5\,\mathrm{K}$ showing a monolayer graphene patch on a Pt(111) surface. Several nanobubbles are visible. Right: STS spectra at $7.5\,\mathrm{K}$ showing locations of $n = 0$, $n = 1$, and $n = 2$ pseudo-Landau levels. Lower left: The energies of the pseudo-LLs are proportional to $|n|^{1/2}$, in agreement with the Dirac theory.

and therefore the effective strain gauge field components are

$$\begin{aligned} Q_x &= -\frac{1}{2\sqrt{3}\,a}\,(3\varepsilon_1 + \varepsilon_2) = -2\eta\,Y/\lambda \\ Q_y &= \frac{1}{2a}\,(\varepsilon_2 - \varepsilon_1) = 2\eta X/\lambda \quad . \end{aligned} \tag{1.124}$$

This is wonderful! The effective strain gauge field corresponds to a uniform fictitious magnetic field, with

$$\partial_x Q_y - \partial_y Q_x = 4\eta/\lambda \quad . \tag{1.125}$$

### 1.4.6 One-dimensional analog

There is a one-dimensional version of the strain gauge field. Consider a bipartite one-dimensional chain with alternating hoppings $t_1 \equiv ut$ and $t_2 \equiv vt$. The Hamiltonian at wavevector $k$ is

$$H_k = -t \begin{pmatrix} 0 & \gamma_k \\ \gamma_k^* & 0 \end{pmatrix} \qquad \text{where} \qquad \gamma_k = u + v\,e^{ik} \quad . \tag{1.126}$$

The energy eigenvalues are $\pm t|\gamma_k|$ . For uniform hopping $u = v = 1$, the gap collapses at $k = \pi$ . Writing $u = 1 + \frac{1}{2}\varepsilon$ and $v = 1 - \frac{1}{2}\varepsilon$ , and $k = \pi + q$, we obtain, to lowest order and in the long wavelength limit,

$$\pi^\dagger \equiv \gamma_{\pi+q} = -iq + \varepsilon = -\partial_x + \varepsilon(x) \quad . \tag{1.127}$$

Note $[\pi, \pi^\dagger] = 2\,\varepsilon'(x) \equiv \pm r^2$ . Assuming $r$ is both constant and real, we then define $\pi \equiv r\,\alpha$, and we once again have $H_+ = rt \begin{pmatrix} 0 & \alpha \\ \alpha^\dagger & 0 \end{pmatrix}$ and $H_- = rt \begin{pmatrix} 0 & \alpha^\dagger \\ \alpha & 0 \end{pmatrix}$, with the Dirac spectrum $E_0 = 0$ and $E_{\pm,n} = \sqrt{n}\,r\,t$ .

## 1.5 An Electron on a Torus

### 1.5.1 Constraints of finite geometry

On an infinite plane, each Landau level is infinitely degenerate, with the number of states per Landau level per unit area given by $B/\phi_0 = 1/2\pi\ell^2$. As diagonalizing infinite-dimensional matrices is time-consuming, it is useful to impose a finite geometry, rendering each Landau level finite in size. Computational approaches have been exceedingly important in the development of the theory of the FQHE, often providing essential insights. In principle one could choose any orientable two-dimensional manifold as a base space, but the two simplest and most useful geometries have been the sphere and the torus. Use of the spherical geometry was pioneered by Duncan Haldane[21]. Here we will focus on the torus geometry[22]. One reason the torus is particularly useful is that it is not simply connected. Topologically, it is a product of two circles: $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$. This furnishes us with an opportunity to impose generalized periodic boundary conditions, which physically corresponds to threading each of the two circles with flux. Differentiating with respect to these fluxes can tell us about the Hall conductivity! So when it comes to finite geometries, to paraphrase Orwell: *sphere good, torus better!*

On the torus, we require that the area $\Omega = \hat{\boldsymbol{z}} \cdot \boldsymbol{L}_1 \times \boldsymbol{L}_2 = 2\pi\ell^2 N_\phi$ be quantized in units of $2\pi\ell^2$, where $N_\phi = B\Omega/\phi_0$ is the number of flux quanta, if we are to have $\big[\, t(\boldsymbol{L}_1)\,,\, t(\boldsymbol{L}_2)\,\big] = 0$. Then since $V(\boldsymbol{r} + \boldsymbol{L}_a) = V(\boldsymbol{r})$, we have that $\big\{\, H\,,\, t(\boldsymbol{L}_1)\,,\, t(\boldsymbol{L}_2)\,\big\}$ is a complete set of commuting observables. Since the $t(\boldsymbol{L}_{1,2})$ are unitary, the eigenstates $|\,\psi_\alpha\,\rangle$ of $H$ may be chosen to satisfy

$$t(\boldsymbol{L}_a)\,|\,\psi_\alpha\,\rangle = e^{i\theta_a}\,|\,\psi_\alpha\,\rangle \tag{1.128}$$

for all $\alpha \in \big\{1, \ldots, N_\phi\big\}$. The dimension of the Hilbert space is $N_\phi$, as we shall see.

We will consider potentials $V(\boldsymbol{r})$ which are periodic on the torus. This is equivalent to a periodic potential, and any reciprocal lattice vector $\boldsymbol{G}$ as

$$\boldsymbol{G} = n_1\,\boldsymbol{b}_1 + n_2\,\boldsymbol{b}_2 \quad , \tag{1.129}$$

[21] See F. D. M. Haldane, *Phys. Rev. Lett.* **31**, 605 (1983).
[22] Also largely pioneered by Haldane.

where

$$\boldsymbol{b}_1 = \frac{2\pi}{\Omega}\,\boldsymbol{L}_2 \times \hat{\boldsymbol{z}} \qquad , \qquad \boldsymbol{b}_2 = \frac{2\pi}{\Omega}\,\hat{\boldsymbol{z}} \times \boldsymbol{L}_1 \quad , \tag{1.130}$$

with integer $n_{1,2}$, where $\Omega = \hat{\boldsymbol{z}} \cdot \boldsymbol{L}_1 \times \boldsymbol{L}_2 = 2\pi\ell^2 N_\phi$ is the area of the torus. With the above definitions of the primitive reciprocal lattice vectors $\boldsymbol{b}_{1,2}$, we have $\boldsymbol{b}_a \cdot \boldsymbol{L}_{a'} = 2\pi\delta_{aa'}$. Thus,

$$\ell^2 \hat{\boldsymbol{z}} \times \boldsymbol{G} = \frac{2\pi\ell^2}{\Omega}\left(n_1\,\boldsymbol{L}_2 - n_2\,\boldsymbol{L}_1\right) \quad . \tag{1.131}$$

Thus,

$$\left(\ell^2 \hat{\boldsymbol{z}} \times \boldsymbol{G}\right) \times \boldsymbol{L}_a = -2\pi\ell^2\, n_a\, \hat{\boldsymbol{z}} \quad , \tag{1.132}$$

and therefore $\left[\, t\big(\ell^2 \hat{\boldsymbol{z}} \times \boldsymbol{G}\big)\,,\, t(\boldsymbol{L}_a)\,\right] = 0$ for $a = 1, 2$.

For any vector $\boldsymbol{Q}$, we define its complexification as $Q \equiv Q_x + iQ_y$. The complexified elementary reciprocal lattice vectors are then

$$b_1 = b_{1,x} + ib_{1,y} = -\frac{2\pi i}{\Omega} L_2 \qquad , \qquad b_2 = b_{2,x} + ib_{2,y} = \frac{2\pi i}{\Omega} L_1 \quad . \tag{1.133}$$

The modular parameter $\tau = \tau_1 + i\tau_2$ is defined as the complex ratio

$$\tau \equiv \frac{L_2}{L_1} = \frac{L_{2,x} + iL_{2,y}}{L_{1,x} + iL_{1,y}} \quad . \tag{1.134}$$

For a general reciprocal lattice vector $\boldsymbol{G} = n_1\boldsymbol{b}_1 + n_2\boldsymbol{b}_2$, then, we have

$$G = G_x + iG_y = \frac{2\pi i}{\Omega}\, L_1\left(n_2 - n_1\tau\right) \quad . \tag{1.135}$$

The unit cell area is then

$$\Omega = 2\pi\ell^2 N = \mathsf{Im}\left(\bar{L}_1 L_2\right) = |L_1|^2\,\tau_2 \quad . \tag{1.136}$$

Then we have

$$\tfrac{1}{4}\boldsymbol{G}^2\ell^2 = \tfrac{1}{4}|G|^2\ell^2 = \frac{\pi}{2N_\phi\tau_2}\Big((n_1\tau_1 - n_2)^2 + n_1^2\,\tau_2^2\Big) \quad . \tag{1.137}$$

## 1.5.2 Lowest Landau level Hamiltonian

The potential may be written in terms of its Fourier components, *viz.*

$$\begin{aligned} V(\boldsymbol{r}) &= \sum_{\boldsymbol{G}} V_{\boldsymbol{G}}\, e^{i\boldsymbol{G}\cdot\boldsymbol{r}} = \sum_{\boldsymbol{G}} V_{\boldsymbol{G}}\, e^{i(G\bar{z} + \bar{G}z)/2} \\ &= \sum_{\boldsymbol{G}} V_{\boldsymbol{G}}\, e^{\ell^2(G\pi^\dagger - \bar{G}\pi)/2\hbar}\, e^{\ell^2(\bar{G}\kappa - G\kappa^\dagger)/2\hbar} \\ &= \sum_{\boldsymbol{G}} V_{\boldsymbol{G}}\, e^{-\boldsymbol{G}^2\ell^2/4}\, e^{\ell^2 G\pi^\dagger/2\hbar} e^{-\ell^2\bar{G}\pi/2\hbar}\, t\big(\ell^2\hat{\boldsymbol{z}} \times \boldsymbol{G}\big) \quad . \end{aligned} \tag{1.138}$$

If we project onto the lowest Landau level, we obtain

$$\widetilde{V} = \mathrm{P}_0\, V(\boldsymbol{r})\, \mathrm{P}_0 = \sum_{\boldsymbol{G}} V_{\boldsymbol{G}}\, e^{-\boldsymbol{G}^2\ell^2/4}\, t\big(\ell^2 \hat{\boldsymbol{z}} \times \boldsymbol{G}\big) \quad . \tag{1.139}$$

Define the unitary operators

$$t_1 \equiv t\big(\boldsymbol{L}_1/N_\phi\big) \qquad , \qquad t_2 \equiv t\big(\boldsymbol{L}_2/N_\phi\big) \quad . \tag{1.140}$$

Then it is easy to show

$$t_1\, t_2 = e^{2\pi i/N_\phi}\, t_2\, t_1 \quad . \tag{1.141}$$

Furthermore, we have

$$t\big(\ell^2 \hat{\boldsymbol{z}} \times \boldsymbol{G}\big) = t\left(\frac{n_1\, \boldsymbol{L}_2}{N_\phi} - \frac{n_2\, \boldsymbol{L}_1}{N_\phi}\right) = e^{-i\pi n_1 n_2/N_\phi}\, t_2^{n_1}\, t_1^{-n_2} \quad . \tag{1.142}$$

We can define an $N_\phi$-element basis $\big\{ |\, k\,\rangle \big\}$ which satisfies the following:

$$\begin{aligned} t_1\, |\, k\,\rangle &= e^{i\theta_1/N_\phi}\, |\, k-1\,\rangle \\ t_2\, |\, k\,\rangle &= e^{i\theta_2/N_\phi}\, e^{2\pi i k/N_\phi}\, |\, k\,\rangle \quad , \end{aligned} \tag{1.143}$$

with $|\, k + N_\phi\,\rangle \equiv |\, k\,\rangle$. Note that $t_1\, t_2\, |\, k\,\rangle = e^{2\pi i/N_\phi}\, t_2\, t_1\, |\, k\,\rangle$ for all $k$, and furthermore that $t(\boldsymbol{L}_a)\, |\, k\,\rangle = t_a^{N_\phi}\, |\, k\,\rangle = e^{i\theta_a}\, |\, k\,\rangle$ for all allowed $a$ and $k$. Thus,

$$\begin{aligned} \langle\, k\,|\, t\big(\ell^2 \hat{\boldsymbol{z}} \times \boldsymbol{G}\big)\,|\, k'\,\rangle &= e^{-i\pi n_1 n_2/N_\phi}\, e^{in_1\theta_2/N_\phi}\, e^{2\pi i k n_1/N_\phi}\, e^{-in_2\theta_1/N_\phi}\, \langle\, k\,|\, k' + n_2\,\rangle \\ &= e^{-i\pi n_1 n_2/N_\phi}\, e^{i(n_1\theta_2 - n_2\theta_1)/N_\phi}\, e^{2\pi i k n_1/N_\phi}\, \tilde{\delta}_{k,k'+n_2} \quad , \end{aligned} \tag{1.144}$$

where $\tilde{\delta}_{k,l} \equiv \delta_{k,l \bmod N_\phi}$. Thus, the Hamiltonian for our system is

$$H_{kk'}(\theta_1,\theta_2) = \sum_{n_1,n_2} V_{n_1,n_2}\, e^{-\pi\left[(n_1\tau_1 - n_2)^2 + n_1^2\tau_2^2\right]/2N_\phi\tau_2}\, e^{-i\pi n_1 n_2/N_\phi}\, e^{i(n_1\theta_2 - n_2\theta_1)/N_\phi}\, e^{2\pi i n_1 k/N_\phi}\, \tilde{\delta}_{k,k'+n_2} \quad . \tag{1.145}$$

Checking that the Hamiltonian is Hermitian, we have

$$\begin{aligned} H^*_{kk'} &= \sum_{n_1,n_2} V^*_{n_1,n_2}\, e^{-\pi(n_1\tau_1 - n_2)^2/2N_\phi\tau_2}\, e^{-\pi n_1^2\tau_2/2N_\phi}\, e^{i\pi n_1 n_2/N_\phi}\, e^{-i(n_1\theta_2 - n_2\theta_1)/N_\phi}\, e^{-2\pi i n_1 k/N_\phi}\, \tilde{\delta}_{k,k'+n_2} \\ &= \sum_{n_1,n_2} V_{n_1,n_2}\, e^{-\pi(n_1\tau_1 - n_2)^2/2N_\phi\tau_2}\, e^{-\pi n_1^2\tau_2/2N_\phi}\, e^{i\pi n_1 n_2/N_\phi}\, e^{i(n_1\theta_2 - n_2\theta_1)/N_\phi}\, e^{2\pi i n_1 (k'-n_2)/N_\phi}\, \tilde{\delta}_{k',k+n_2} \\ &= \sum_{n_1,n_2} V_{n_1,n_2}\, e^{-\pi(n_1\tau_1 - n_2)^2/2N_\phi\tau_2}\, e^{-\pi n_1^2\tau_2/2N_\phi}\, e^{-i\pi n_1 n_2/N_\phi}\, e^{i(n_1\theta_2 - n_2\theta_1)/N_\phi}\, e^{2\pi i n_1 k'/N_\phi}\, \tilde{\delta}_{k',k+n_2} = H_{k'k} \quad , \end{aligned}$$

where we have used $V^*_{n_1,n_2} = V_{-n_1,-n_2}$ and then replaced $n_{1,2}$ with $-n_{1,2}$ in the second line.

As an example, consider a case where $V(\boldsymbol{r}) = V_{1,1}\, e^{i(\boldsymbol{b}_1+\boldsymbol{b}_2)\cdot\boldsymbol{r}} + V^*_{1,1}\, e^{-i(\boldsymbol{b}_1+\boldsymbol{b}_2)\cdot\boldsymbol{r}}$ and $N_\phi = 3$. We then find

$$H = \begin{pmatrix} 0 & -V^*_{1,1}\, e^{i\alpha} & V_{1,1}\, e^{i\pi/3}\, e^{-i\alpha} \\ -V_{1,1}\, e^{-i\alpha} & 0 & V^*_{1,1}\, e^{i\pi/3}\, e^{i\alpha} \\ V^*_{1,1}\, e^{-i\pi/3}\, e^{i\alpha} & V_{1,1}\, e^{-i\pi/3}\, e^{-i\alpha} & 0 \end{pmatrix} e^{-\pi\left[(\tau_1-1)^2+\tau_2^2\right]/6\tau_2} \quad , \tag{1.146}$$

where $\alpha \equiv \frac{1}{3}(\theta_1 - \theta_2)$. For example, to compute $H_{kk'}$ with $k = 1$ and $k' = 2$, we need $n_2 = -1$ to satisfy the Kronecker delta in eqn. 1.145, and therefore $n_1 = -1$ as well, corresponding to $V_{-1,-1} = V^*_{1,1}$. Working out the phase of the matrix element, we then have $e^{-i\pi/3}\, e^{i(\theta_1-\theta_2)/3}\, e^{-2\pi i/3} = -e^{i\alpha}$. For $k = 1$ and $k' = 3$, we need $n_2 = 1$ in order to satisfy $k = k' + 1 \bmod N$. Thus $n_1 = 1$ as well, corresponding to $V_{1,1}$, and the phase of the matrix element is $e^{-i\pi/3}\, e^{i(\theta_2-\theta_1)/3}\, e^{2\pi i/3} = e^{i\pi/3}\, e^{-i\alpha}$.

A detailed discussion of the LLL wavefunctions on the torus is given in §1.8 below.

## 1.6 Lattice Models and Hofstadter's Butterfly

### 1.6.1 Tight binding with $B = 0$

As you may have heard, solids are composed of atoms[23]. Suppose we have an orthonormal set of orbitals $|\, a\boldsymbol{R}\,\rangle$, where $a$ labels the orbital and $\boldsymbol{R}$ denotes a Bravais lattice site. The label $a$ may refer to different orbitals associated with the atom at $\boldsymbol{R}$, or it may label orbitals on other atoms in the unit cell defined by $\boldsymbol{R}$. The most general tight binding Hamiltonian we can write is

$$H = \sum_{\boldsymbol{R},\boldsymbol{R}'} \sum_{a,a'} H_{aa'}(\boldsymbol{R}-\boldsymbol{R}')\, |\, a\boldsymbol{R}\,\rangle\langle\, a'\boldsymbol{R}'\,| \quad , \tag{1.147}$$

where $H_{aa'}(\boldsymbol{R}-\boldsymbol{R}') = H^*_{a'a}(\boldsymbol{R}'-\boldsymbol{R}) = \langle\, a, \boldsymbol{R}\,|\, H\,|\, a', \boldsymbol{R}'\,\rangle$ is the Hamiltonian matrix, whose rows and columns are indexed by a composite index combining both the unit cell label $\boldsymbol{R}$ and the orbital label $a$. When $\boldsymbol{R} = \boldsymbol{R}'$ and $a = a'$, the term $H_{aa}(0) = \varepsilon_a$ is the energy of a single electron in an isolated $a$ orbital. For all other cases, $H_{aa'}(\boldsymbol{R}-\boldsymbol{R}') = -t_{aa'}(\boldsymbol{R}-\boldsymbol{R}')$ is the hopping integral between the $a$ orbital in unit cell $\boldsymbol{R}$ and the $a'$ orbital in unit cell $\boldsymbol{R}'$. Let's write an eigenstate $|\,\psi\,\rangle$ as

$$|\,\psi\,\rangle = \sum_{\boldsymbol{R}} \sum_{a} \psi_{a\boldsymbol{R}}\, |\, a\boldsymbol{R}\,\rangle \quad . \tag{1.148}$$

Applying the Hamiltonian to $|\,\psi\,\rangle$, we obtain the coupled equations

$$\sum_{\boldsymbol{R},\boldsymbol{R}'} \sum_{a,a'} H_{aa'}(\boldsymbol{R}-\boldsymbol{R}')\, \psi_{a'\boldsymbol{R}'}\, |\, a\boldsymbol{R}\,\rangle = E \sum_{\boldsymbol{R}} \sum_{a} \psi_{a\boldsymbol{R}}\, |\, a\boldsymbol{R}\,\rangle \quad . \tag{1.149}$$

[23] A somewhat more nuanced description: solids are composed of ions and electrons.

Since the $|\, a\boldsymbol{R}\,\rangle$ basis is complete, we must have that the coefficients of $|\, a\boldsymbol{R}\,\rangle$ on each side agree. Therefore,

$$\sum_{\boldsymbol{R}'}\sum_{a'} H_{aa'}(\boldsymbol{R}-\boldsymbol{R}')\,\psi_{a'\boldsymbol{R}'} = E\,\psi_{a\boldsymbol{R}} \quad . \tag{1.150}$$

**Bloch's theorem**

We now use Bloch's theorem, which says that each eigenstate may be labeled by a wavevector $\boldsymbol{k}$, with $\psi_{a\boldsymbol{R}} = \frac{1}{\sqrt{N}}\,u_a(\boldsymbol{k})\,e^{i\boldsymbol{k}\cdot\boldsymbol{R}}$. The $N^{-1/2}$ prefactor is a normalization term. Multiplying each side by $e^{-i\boldsymbol{k}\cdot\boldsymbol{R}}$, we have

$$\sum_{a'}\left(\sum_{\boldsymbol{R}'} H_{aa'}(\boldsymbol{R}-\boldsymbol{R}')\,e^{-i\boldsymbol{k}\cdot(\boldsymbol{R}-\boldsymbol{R}')}\right) u_{a'\boldsymbol{k}} = E(\boldsymbol{k})\,u_{a\boldsymbol{k}} \quad , \tag{1.151}$$

which may be written as

$$\sum_{a'} \hat{H}_{aa'}(\boldsymbol{k})\,u_{a'}(\boldsymbol{k}) = E(\boldsymbol{k})\,u_a(\boldsymbol{k}) \quad , \tag{1.152}$$

where

$$\hat{H}_{aa'}(\boldsymbol{k}) = \sum_{\boldsymbol{R}} H_{aa'}(\boldsymbol{R})\,e^{-i\boldsymbol{k}\cdot\boldsymbol{R}} \quad . \tag{1.153}$$

Thus, for each crystal wavevector $\boldsymbol{k}$, the $u_{a\boldsymbol{k}}$ are the eigenfunctions of the $r\times r$ Hermitian matrix $\hat{H}_{aa'}(\boldsymbol{k})$. The energy eigenvalues at wavevector $\boldsymbol{k}$ are given by $\mathrm{spec}\big(\hat{H}(\boldsymbol{k})\big)$, *i.e.* by the set of eigenvalues of the matrix $\hat{H}(\boldsymbol{k})$. There are $r$ such solutions (some of which may be degenerate), which we distinguish with a band index $n$, and we denote $u_{na}(\boldsymbol{k})$ and $E_n(\boldsymbol{k})$ as the corresponding eigenvectors and eigenvalues. We sometimes will use the definition $\hat{t}_{aa'}(\boldsymbol{k}) \equiv -\hat{H}_{aa'}(\boldsymbol{k})$ for the matrix of hopping integrals.

In Eqn. 1.147, $\boldsymbol{R}$ and $\boldsymbol{R}'$ label Bravais lattice sites, while $a$ and $a'$ label orbitals. We stress that these orbitals don't necessarily have to be located on the same ion. We should think of $\boldsymbol{R}$ and $\boldsymbol{R}'$ labeling *unit cells*, each of which is indeed associated with a Bravais lattice site. For example, in the case of graphene, $|\, a\boldsymbol{R}\,\rangle$ represents an orbital on the $a$ sublattice in unit cell $\boldsymbol{R}$. The eigenvalue equation may be written

$$\hat{H}_{aa'}(\boldsymbol{k})\,u_{na'}(\boldsymbol{k}) = E_n(\boldsymbol{k})\,u_{na}(\boldsymbol{k}) \quad , \tag{1.154}$$

where $n$ is the band index. The function $u_{na}(\boldsymbol{k})$ is the *internal wavefunction* within a given cell, and corresponds to the cell function $u_{n\boldsymbol{k}}(\boldsymbol{r})$ in the continuum, with $a \leftrightarrow (\boldsymbol{r}-\boldsymbol{R})$ labeling a position within each unit cell. The full Bloch state may then be written

$$|\,\psi_{n\boldsymbol{k}}\,\rangle = |\,\boldsymbol{k}\,\rangle \otimes |\,u_{n\boldsymbol{k}}\,\rangle \quad , \tag{1.155}$$

so that

$$\begin{aligned}\psi_{n\boldsymbol{k}}(\boldsymbol{R},a) &= \Big(\langle\,\boldsymbol{R}\,|\otimes\langle\,a\,|\Big)\Big(|\,\boldsymbol{k}\,\rangle\otimes|\,u_{n\boldsymbol{k}}\,\rangle\Big)\\ &= \langle\,\boldsymbol{R}\,|\,\boldsymbol{k}\,\rangle\,\langle\,a\,|\,u_{n\boldsymbol{k}}\,\rangle = \frac{1}{\sqrt{N}}\,e^{i\boldsymbol{k}\cdot\boldsymbol{R}}\,u_{na}(\boldsymbol{k}) \quad .\end{aligned} \tag{1.156}$$

Here we have chosen a normalization $\sum_a \left|u_{na}(\boldsymbol{k})\right|^2 = 1$ within each unit cell, which entails the overall normalization $\sum_{\boldsymbol{R},a}\left|\psi_{n\boldsymbol{k}}(\boldsymbol{R},a)\right|^2 = 1$.

### 1.6.2 Go flux yourself : how to add magnetic fields

To simplify matters, we consider only $s$-orbitals on two-dimensional lattices. The general tight-binding Hamiltonian is written

$$H = -\sum_{\boldsymbol{r}<\boldsymbol{r}'}\Big(t_{\boldsymbol{r}\boldsymbol{r}'}\,c^\dagger_{\boldsymbol{r}}\,c_{\boldsymbol{r}'} + t^*_{\boldsymbol{r}\boldsymbol{r}'}\,c^\dagger_{\boldsymbol{r}'}\,c_{\boldsymbol{r}}\Big) \quad , \tag{1.157}$$

where the notation $\boldsymbol{r} < \boldsymbol{r}'$ means that each pair $(\boldsymbol{r},\boldsymbol{r}')$ summed only once. We may write $t_{\boldsymbol{r}\boldsymbol{r}'} = t^*_{\boldsymbol{r}'\boldsymbol{r}} = |t_{\boldsymbol{r}\boldsymbol{r}'}|\exp(iA_{\boldsymbol{r}\boldsymbol{r}'})$, where $A_{\boldsymbol{r}\boldsymbol{r}'}$ is a gauge field living on the links of the lattice. Let $p$ denote a plaquette on the lattice. Then the dimensionless flux $\phi_p$ (in units of $\hbar c/e$) through plaquette $p$ is

$$\phi_p = \sum_{\langle\boldsymbol{r}\boldsymbol{r}'\rangle\in\partial p} A_{\boldsymbol{r}\boldsymbol{r}'} \quad , \tag{1.158}$$

where the sum is taken in a counterclockwise fashion along the links on the boundary of $p$. The tight-binding Hamiltonian exhibits a *gauge invariance* under

$$\begin{aligned}c_{\boldsymbol{r}} &\to e^{i\alpha_{\boldsymbol{r}}}\,c_{\boldsymbol{r}}\\ t_{\boldsymbol{r}\boldsymbol{r}'} &\to e^{i(\alpha_{\boldsymbol{r}}-\alpha_{\boldsymbol{r}'})}\,t_{\boldsymbol{r}\boldsymbol{r}'} \quad .\end{aligned} \tag{1.159}$$

Consider now the case of the square lattice. It is clear that any configuration of the $A_{\boldsymbol{r}\boldsymbol{r}'}$ which is periodic in the structural unit cell, *i.e.* under translations by elementary direct lattice vectors, must correspond to $\phi_p = 0$ for every plaquette $p$[24]. This is because the phase $A_{\boldsymbol{r}\boldsymbol{r}'}$ is associated with the *directed link* from $\boldsymbol{r}$ to $\boldsymbol{r}'$, and parallel links on opposite sides of the elementary square plaquette will yield equal and opposite values of $A_{\boldsymbol{r}\boldsymbol{r}'}$ because they are traversed in opposite directions. *In order to describe nonzero flux per plaquette, the configuration of the lattice gauge field*

[24]More precisely, if $A_{\boldsymbol{r}\boldsymbol{r}'}$ is periodic in the structural unit cell, then each structural unit cell is congruent to a zero flux state. However, it may be that a structural cell is comprised of more than one elementary plaquette, as is the case with the triangular lattice (each structural cell consists of two triangles), or that there are closed loops which don't correspond to a structural unit cell due to further neighbor hoppings. In such cases, there may be closed loops on the lattice whose flux is not congruent to zero. See §1.6.3 for some examples.

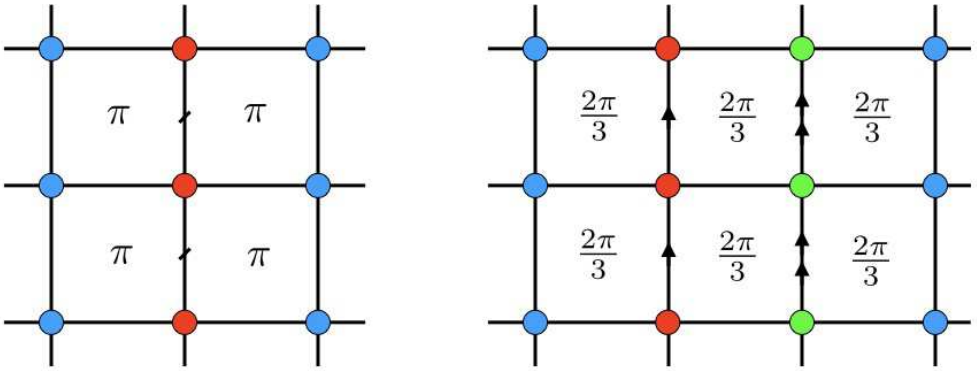


Figure 1.16: Gauges for the square lattice Hofstadter model. Left: $\phi = \pi$ case. $t_{\boldsymbol{r}\boldsymbol{r}'} = t$ on all links except those with slashes, where $t_{\boldsymbol{r}\boldsymbol{r}'} = -t$. Right: $\phi = \frac{2}{3}\,\pi$. Each arrow corresponds to a factor of $\exp(2\pi i/3)$.

*$A_{\boldsymbol{r}\boldsymbol{r}'}$ must break lattice translational symmetry*[25]. Consider the case where $\phi = \pi$ in each plaquette. A configuration for the gauge field $A_{\boldsymbol{r}\boldsymbol{r}'}$ yielding this flux distribution is shown in the left panel of Fig. 1.16. All links have $A_{\boldsymbol{r}\boldsymbol{r}'} = 0$, hence $t_{\boldsymbol{r}\boldsymbol{r}'} = t\exp(iA_{\boldsymbol{r}\boldsymbol{r}'}) = t$, except for the links depicted with slashes, for which $A_{\boldsymbol{r}\boldsymbol{r}'} = \pi$ and $t_{\boldsymbol{r}\boldsymbol{r}'} = -t$. The *magnetic unit cell* is now a $2 \times 1$ block consisting of one cell from each sublattice (blue and red). We call this a magnetic unit cell to distinguish it from the *structural unit cell* of the underlying square lattice. The *structural* Bravais lattice is square, with elementary direct lattice vectors are $\boldsymbol{a}_1 = a\hat{\boldsymbol{x}}$ and $\boldsymbol{a}_2 = a\hat{\boldsymbol{y}}$. But the *magnetic* Bravais lattice is rectangular, with elementary RLVs $\boldsymbol{a}_1 = 2a\hat{\boldsymbol{x}}$ and $\boldsymbol{a}_2 = a\hat{\boldsymbol{y}}$. From Bloch's theorem, the phase of the wavefunction varies by $\exp(i\boldsymbol{k}\cdot\boldsymbol{a}_1) \equiv \exp(i\theta_1)$ across the unit cell in the $x$-direction, and by $\exp(i\boldsymbol{k}\cdot\boldsymbol{a}_2) \equiv \exp(i\theta_2)$ in the $y$-direction. The Hamiltonian is

$$\hat{H}(\boldsymbol{\theta}) = -t\begin{pmatrix} 2\cos\theta_2 & 1+e^{-i\theta_1} \\ 1+e^{i\theta_1} & -2\cos\theta_2 \end{pmatrix} \tag{1.160}$$

The energy eigenvalues are $E_\pm(\boldsymbol{\theta}) = \pm 2t\sqrt{\cos^2(\frac{1}{2}\theta_1) + \cos^2\theta_2}$. The band gap collapses at two points: $(\theta_1, \theta_2) = (\pi, \pm\frac{1}{2}\pi)$. Writing $(\theta_1, \theta_2) = (\pi + \delta_1\,,\, \pm\frac{1}{2}\pi + \delta_2)$, we find

$$E_\pm(\boldsymbol{\theta}) = \pm 2t\sqrt{\sin^2(\tfrac{1}{2}\delta_1) + \sin^2\delta_2} = \pm 2ta\,\sqrt{\tfrac{1}{4}q_1^2 + q_2^2} \quad , \tag{1.161}$$

which is a Dirac cone! Thus, the dispersion for the square lattice $\pi$ flux model has two Dirac points. Here, $\boldsymbol{q} = \boldsymbol{k} - \boldsymbol{k}_{\rm D}$ is the wavevector measured from either Dirac point.

The $\pi$ flux state is time-reversal symmetric, since under time reversal we have $\exp(iA_{\boldsymbol{r}\boldsymbol{r}'}) \to \exp(-iA_{\boldsymbol{r}\boldsymbol{r}'})$, hence $\phi_p \to -\phi_p$. But flux is only defined modulo $2\pi$, hence $\pi \to -\pi \cong \pi$ yields the same flux configuration.

A more interesting state of affairs pertains for the case $\phi = \frac{2}{3}\pi$, for which a valid gauge configuration $A_{\boldsymbol{r}\boldsymbol{r}'}$ is shown in the right panel of Fig. 1.16. Now there are three sites per unit

[25] By "nonzero" flux, we mean $\phi \bmod 2\pi \neq 0$.

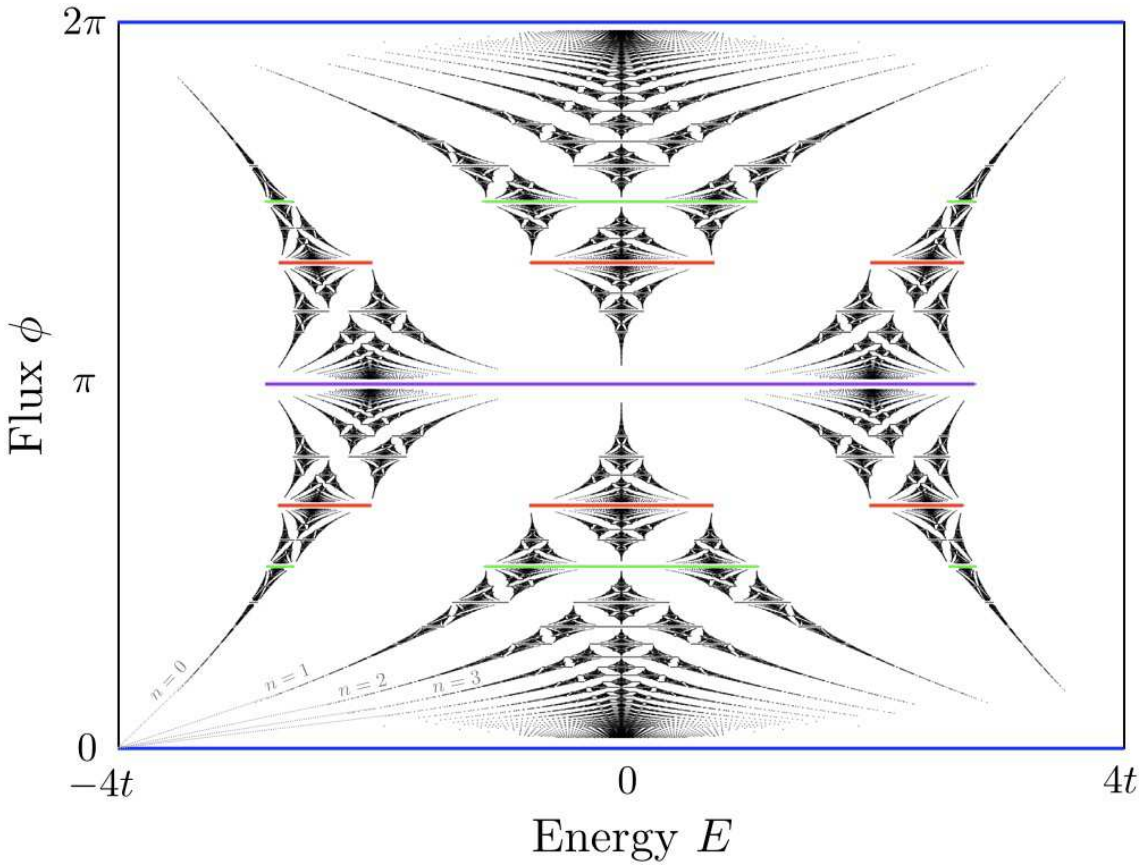


Figure 1.17: Magnetic subbands for the square lattice Hofstadter model for flux per plaquette $\phi \in [0, 2\pi]$. Blue bands at $\phi = 0$ and $\phi = 2\pi$ have the full bandwidth $W = 8t$. At $\phi = \pi$ (purple), there are two subbands with $E_- \in [-2\sqrt{2}t, 0]$ and $E_+ \in [0, 2\sqrt{2}t]$ which touch at $E = 0$. Similarly, at $\phi = \pm\frac{1}{2}\pi$ (green), there are four subbands, with the central two bands!bagain touching at $E = 0$. At $\phi = \pm\frac{2}{3}\pi$ (red), there are three subbands. Continuum Landau levels are shown radiating from the lower left corner.

cell: A (blue), B (red), and C (green). The Bloch phase accrued across the magnetic unit cell in the $\pm\hat{\boldsymbol{x}}$ direction is $e^{\pm i\theta_1}$, and in the $\pm\hat{\boldsymbol{y}}$ direction is $e^{\pm i\theta_2}$. Thus

$$\hat{H}(\boldsymbol{\theta}) = -t \begin{pmatrix} 2\cos\theta_2 & 1 & e^{-i\theta_1} \\ 1 & 2\cos(\theta_2 + \frac{2\pi}{3}) & 1 \\ e^{i\theta_1} & 1 & 2\cos(\theta_2 + \frac{4\pi}{3}) \end{pmatrix} \quad . \tag{1.162}$$

The general case where the flux per structural unit cell is $\phi = 2\pi p/q$ is known as the Hofstadter model[26] In this case, the magnetic unit cell is a $q \times 1$ block, and the resulting $q \times q$ Hamiltonian

[26]See D. R. Hofstadter, *Phys. Rev. B* **14**, 2239 (1976).

is given by

$$\hat{H}(\boldsymbol{\theta}) = -t \begin{pmatrix} 2\cos\theta_2 & 1 & 0 & \cdots & 0 & e^{-i\theta_1} \\ 1 & 2\cos\left(\theta_2 + \frac{2\pi p}{q}\right) & 1 & & & 0 \\ 0 & 1 & 2\cos\left(\theta_2 + \frac{4\pi p}{q}\right) & 1 & & \vdots \\ \vdots & 0 & 1 & \ddots & & \vdots \\ & & & & & 1 \\ e^{i\theta_1} & 0 & & \cdots & 1 & 2\cos\left(\theta_2 + \frac{2\pi(q-1)p}{q}\right) \end{pmatrix} \quad . \tag{1.163}$$

There are thus $q$ magnetic subbands. Note that

$$H(\theta_1, \theta_2 + \tfrac{2\pi p}{q}) = XUH(\theta_1, \theta_2)\, U^\dagger X^\dagger \quad , \tag{1.164}$$

where $X_{ij} = \delta_{i,j+1 \bmod q}$ and $U = \mathsf{diag}(1, e^{i\theta_1}, \cdots, e^{i\theta_1})$. Thus,

$$\mathsf{spec}\, H(\theta_1, \theta_2 + \tfrac{2\pi p}{q}) = \mathsf{spec}\, H(\theta_1, \theta_2) \quad , \tag{1.165}$$

as we saw explicitly in the $q = 2$ case above. A plot of the magnetic subbands in $(E, \phi)$ space, known as *Hofstadter's butterfly*, is shown in Fig. 1.17.

In the limit where the denominator $q$ of the flux $\phi = 2\pi p/q$ is large (for fixed $p$), the flux per cell is very small. We then expect to recover the continuum Landau level spectrum $E_n = (n+\frac{1}{2})\hbar\omega_{\rm c}$. To express this in terms of the flux $\phi$, note that the $B = 0$ dispersion is

$$E(\boldsymbol{k}) = -2t\cos(k_x a) - 2t\cos(k_y a) = -4t + t\boldsymbol{k}^2 a^2 + \ldots \quad , \tag{1.166}$$

which allows us to identify the effective mass $m$ from the coefficient of the $\boldsymbol{k}^2$ term, with the result $m = \hbar^2/2ta^2$. The magnetic field is the flux per unit area, hence $B = \phi\hbar c/ea^2$. Thus,

$$\hbar\omega_{\rm c} = \frac{\hbar eB}{mc} = \frac{\hbar e}{c} \times \frac{\phi\,\hbar c}{ea^2} \times \frac{2ta^2}{\hbar^2} = 2\phi\, t \quad . \tag{1.167}$$

This describes the corners of the Hofstadter butterfly in Fig. 1.17, where continuum Landau levels radiate outward from the energies $\pm 4t$ according to

$$E_n(\phi) = \pm\Big(4t - (2n+1)\,\phi\, t\Big) \qquad \text{and} \qquad E_n(\phi) = \pm\Big(4t - (2n+1)\,(2\pi - \phi)\, t\Big) \quad , \tag{1.168}$$

for $\phi \ll 1$.

### 1.6.3 Unit cells with zero net flux

As mentioned in a footnote above, it is not quite true that a lattice gauge field $A_{rr'}$ which is periodic in the underlying Bravais lattice unit cell leads to zero net flux in every plaquette or

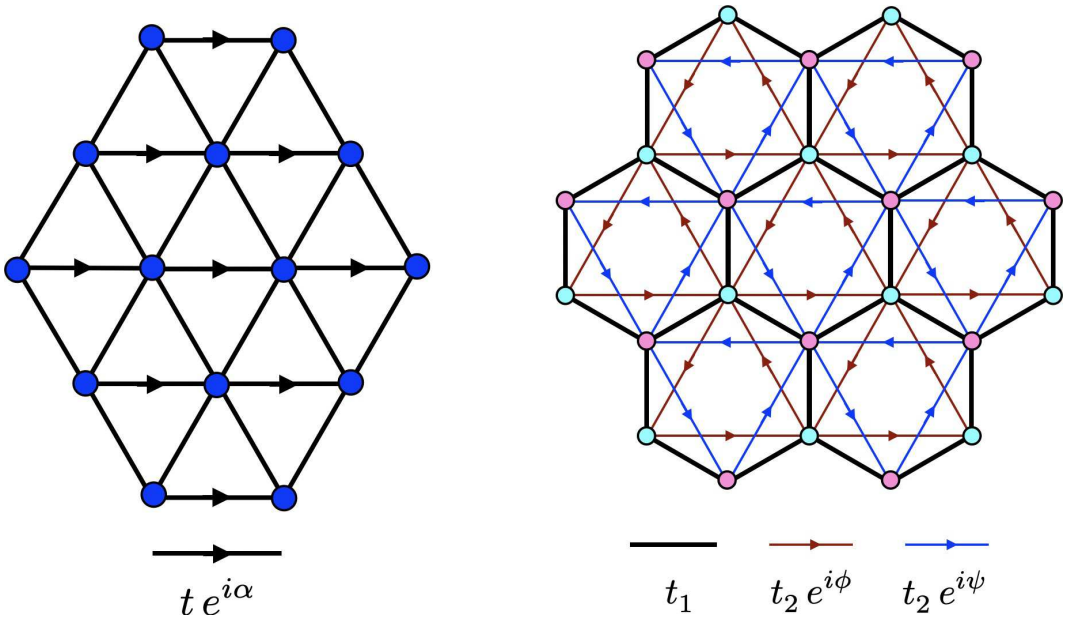


Figure 1.18: Two models with zero net flux per unit cell which still break time reversal symmetry. Left: The unit cell of the triangular lattice consists of two triangles.

closed loop of links on the lattice. Two counterexamples are shown in Fig. 1.18. The first example is that of the triangular lattice, where each structural unit cell is a rhombus consisting of two elementary triangular plaquettes. Consider now the situation where each horizontal link carries a U(1) phase $\alpha$, *i.e.* $A_{\boldsymbol{r}\boldsymbol{r}'} = t\,e^{i\alpha}$, while the remaining links all have $A_{\boldsymbol{r}\boldsymbol{r}'} = 0$. Computing the U(1) flux by taking the directed sum counterclockwise over each triangle, we see that all the up triangles carry flux $\phi_\triangle = \alpha$, while all the down triangles carry flux $\phi_\triangledown = -\alpha \cong 2\pi - \alpha$. If, as before, we take the elementary direct lattice vectors to be $\boldsymbol{a}_{1,2} = a\big(\frac{1}{2}\hat{\boldsymbol{x}} \pm \frac{\sqrt{3}}{2}\hat{\boldsymbol{y}}\big)$ and write $\boldsymbol{k} = \sum_{j=1}^{2} \theta_j\,\boldsymbol{b}_j/2\pi$, with $\boldsymbol{a}_i \cdot \boldsymbol{b}_j = 2\pi\,\delta_{ij}$, then the tight binding Hamiltonian for this triangular lattice model is given by $H = \sum_{\boldsymbol{k}} E_{\boldsymbol{k}}\,a^\dagger_{\boldsymbol{k}}\,a_{\boldsymbol{k}}$, where

$$\begin{aligned} E_{\boldsymbol{k}} &= -2t\cos(\boldsymbol{k}\cdot\boldsymbol{a}_1 + \boldsymbol{k}\cdot\boldsymbol{a}_2 + \alpha) - 2t\cos(\boldsymbol{k}\cdot\boldsymbol{a}_1) - 2t\cos(\boldsymbol{k}\cdot\boldsymbol{a}_2) \\ &= -2t\cos(\theta_1 + \theta_2 + \alpha) - 2t\cos\theta_1 - 2t\cos\theta_2 \quad . \end{aligned} \tag{1.169}$$

A more interesting state of affairs is depicted in the right panel of Fig. 1.18, which is graphene augmented by nearest neighbor same-sublattice hopping terms, which is the celebrated Haldane honeycomb lattice model[27]. Inscribed in each hexagonal cell are one up-triangle of A site, depicted by blue dots in the figure, and one down-triangle of B sites, depicted as pink dots in the figure. Again we take $\boldsymbol{a}_{1,2} = a\big(\frac{1}{2}\hat{\boldsymbol{x}} \pm \frac{\sqrt{3}}{2}\hat{\boldsymbol{y}}\big)$ for the underlying A Bravais lattice, with the basis vectors $\boldsymbol{0}$ and $\boldsymbol{\delta}_1 = a\hat{\boldsymbol{y}}$. The nearest neighbor hoppings between A and B sites all are taken to have amplitude $t_1$, while the inscribed A and B same-sublattice hoppings are taken to be $t_2\,e^{i\phi}$

[27] F. D. M. Haldane, *Phys. Rev. Lett.* **61**, 2015 (1988).

and $t_2\,e^{i\psi}$, respectively, and taken in the counterclockwise direction around the inscribed $\triangle$ and $\triangledown$ paths. An on-site energy term $\pm m$, called the *Semenoff mass*, is added to the hopping terms. One then obtains the tight binding Hamiltonian

$$\begin{aligned} H = \sum_{\boldsymbol{k}} \Big\{ &- t_1 \left(1 + e^{i\boldsymbol{k}\cdot\boldsymbol{a}_1} + e^{-i\boldsymbol{k}\cdot\boldsymbol{a}_2}\right) a^\dagger_{\boldsymbol{k}}\, b_{\boldsymbol{k}} - t_1 \left(1 + e^{-i\boldsymbol{k}\cdot\boldsymbol{a}_1} + e^{i\boldsymbol{k}\cdot\boldsymbol{a}_2}\right) b^\dagger_{\boldsymbol{k}}\, a_{\boldsymbol{k}} \\ &+ \Big[m - 2t_2\,\mathsf{Re}\left(e^{i\phi}\,e^{i\boldsymbol{k}\cdot(\boldsymbol{a}_1+\boldsymbol{a}_2)} + e^{i\phi}\,e^{-i\boldsymbol{k}\cdot\boldsymbol{a}_1} + e^{i\phi}\,e^{-i\boldsymbol{k}\cdot\boldsymbol{a}_2}\right)\Big] a^\dagger_{\boldsymbol{k}}\, a_{\boldsymbol{k}} \\ &\Big[-m - 2t_2\,\mathsf{Re}\left(e^{i\psi}\,e^{-i\boldsymbol{k}\cdot(\boldsymbol{a}_1+\boldsymbol{a}_2)} + e^{i\psi}\,e^{i\boldsymbol{k}\cdot\boldsymbol{a}_1} + e^{i\psi}\,e^{i\boldsymbol{k}\cdot\boldsymbol{a}_2}\right)\Big] b^\dagger_{\boldsymbol{k}}\, b_{\boldsymbol{k}} \Big\} \\ = \sum_{\boldsymbol{k}} &\begin{pmatrix} a^\dagger_{\boldsymbol{k}} & b^\dagger_{\boldsymbol{k}} \end{pmatrix} \begin{pmatrix} H_{\mathsf{AA}}(\boldsymbol{k}) & H_{\mathsf{AB}}(\boldsymbol{k}) \\ H_{\mathsf{BA}}(\boldsymbol{k}) & H_{\mathsf{BB}}(\boldsymbol{k}) \end{pmatrix} \begin{pmatrix} a_{\boldsymbol{k}} \\ b_{\boldsymbol{k}} \end{pmatrix} \quad , \end{aligned} \tag{1.170}$$

where

$$\begin{aligned} H_{\mathsf{AA}}(\boldsymbol{k}) &= m - 2t_2\big[\cos\theta_1 + \cos\theta_2 + \cos(\theta_1+\theta_2)\big]\cos\phi - 2t_2\big[\sin\theta_1 + \sin\theta_2 - \sin(\theta_1+\theta_2)\big]\sin\phi \\ H_{\mathsf{BB}}(\boldsymbol{k}) &= -m - 2t_2\big[\cos\theta_1 + \cos\theta_2 + \cos(\theta_1+\theta_2)\big]\cos\psi + 2t_2\big[\sin\theta_1 + \sin\theta_2 - \sin(\theta_1+\theta_2)\big]\sin\psi \\ H_{\mathsf{AB}}(\boldsymbol{k}) &= H^*_{\mathsf{BA}}(\boldsymbol{k}) = -t_1\big(1 + e^{i\theta_1} + e^{-i\theta_2}\big) \quad . \end{aligned} \tag{1.171}$$

In the Haldane model, $\psi = \phi$, in which case we may write $H(\boldsymbol{k})$ in terms of Pauli matrices, as

$$\begin{aligned} H(\boldsymbol{k}) = &-2t_2\big[\cos\theta_1 + \cos\theta_2 + \cos(\theta_1+\theta_2)\big]\cos\phi - t_1\big(1 + \cos\theta_1 + \cos\theta_2\big)\,\sigma^x \\ &+ t_1\big(\sin\theta_1 - \sin\theta_2\big)\,\sigma^y + \Big(m - 2t_2\big[\sin\theta_1 + \sin\theta_2 - \sin(\theta_1+\theta_2)\big]\sin\phi\Big)\,\sigma^z \quad . \end{aligned} \tag{1.172}$$

What makes the Haldane model so interesting is that its band structure is *topological* over a range of the dimensionless parameters $m/t_2$ and $\phi$. (Without loss of generality, we may set $t_1 \equiv 1$.) More on this below!

### 1.6.4 General flux configuration on the square lattice

More generally, consider a magnetic unit cell formed by an $M \times N$ block of structural unit cells, as depicted in Fig. 1.19. Each structural cell $p$ is labeled by the indices $(m, n)$, where the Bravais lattice site its lower left corner is $ma\hat{\boldsymbol{x}} + na\hat{\boldsymbol{y}}$. To assign the lattice gauge fields, do the following. For $\boldsymbol{r} = ma\hat{\boldsymbol{x}} + na\hat{\boldsymbol{y}}$ and $\boldsymbol{r}' = ma\hat{\boldsymbol{x}} + (n+1)a\hat{\boldsymbol{y}}$ with $n < N$, let $A_{\boldsymbol{r}\boldsymbol{r}'} = \sum_{i=1}^{m-1}\phi_{i,n}$. For the $n = N$, we include the Bloch phase $\theta_2$, so that $A_{\boldsymbol{r}\boldsymbol{r}'} = \theta_2 + \sum_{i=1}^{m-1}\phi_{i,n}$, also noting that $(m, N+1) \cong (m, 1)$. This sets $A_{\boldsymbol{r}\boldsymbol{r}'}$ for all vertical ($y$-directed) links. The only horizontal links for which $A_{\boldsymbol{r}\boldsymbol{r}'}$ are nonzero are those with $\boldsymbol{r} = Ma\hat{\boldsymbol{x}} + na\hat{\boldsymbol{y}}$ and $\boldsymbol{r}' = a\hat{\boldsymbol{x}} + na\hat{\boldsymbol{y}}$; note $(M+1, n) \cong (1, n)$. Then $A_{\boldsymbol{r}\boldsymbol{r}'} = \theta_1 - \sum_{i=1}^{M}\sum_{j=1}^{n}\phi_{i,j}$. One can check that this prescription yields the desired flux configuration, as well as the two Bloch phases.

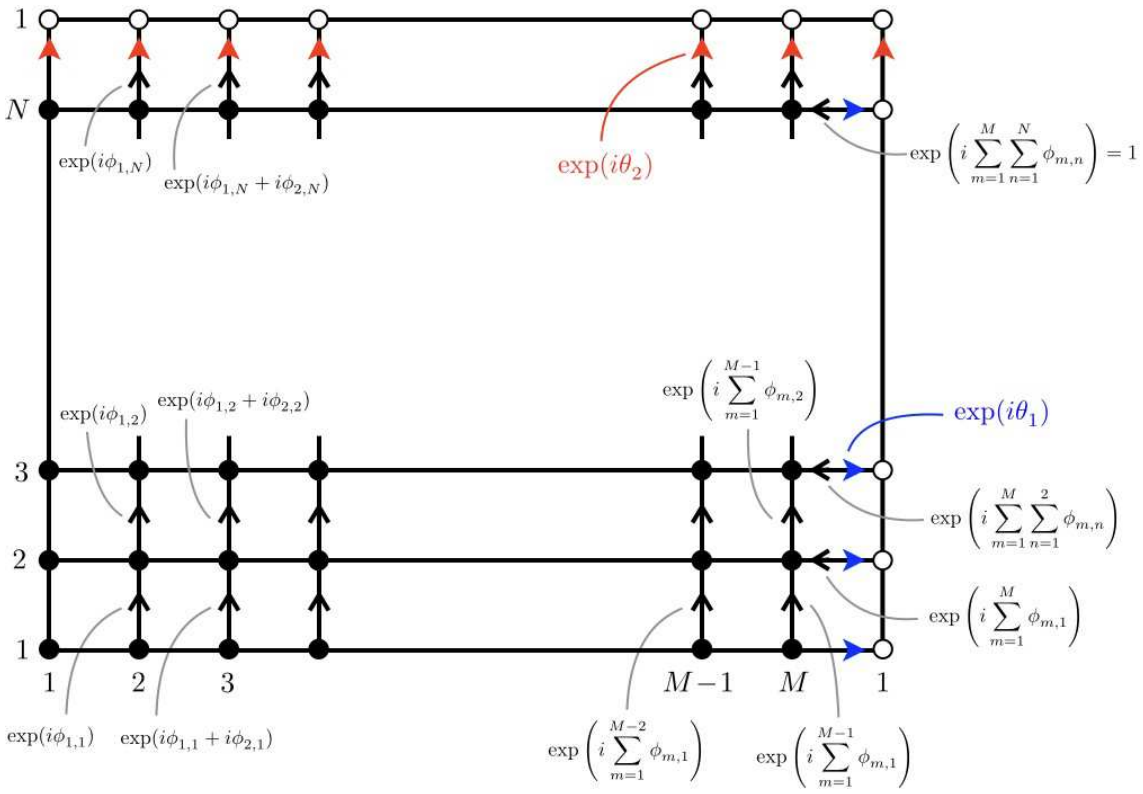


Figure 1.19: Lattice gauge field configuration for a general $M \times N$ rectangular lattice with flux $\phi_{m,n}$ in unit cell $(m, n)$ and Bloch phases $(\theta_1, \theta_2)$.

## 1.7 Berry Phases, Fiber Bundles, Chern Numbers, and TKNN

### 1.7.1 The adiabatic theorem and Berry's phase

Consider a Hamiltonian $H(\boldsymbol{\lambda})$ dependent on a set of parameters $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_K\}$, and let $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle$ satisfy the time-independent Schrödinger equation,

$$H(\boldsymbol{\lambda})\,|\,\varphi_n(\boldsymbol{\lambda})\,\rangle = E_n(\boldsymbol{\lambda})\,|\,\varphi_n(\boldsymbol{\lambda})\,\rangle \quad . \tag{1.173}$$

Now let $\boldsymbol{\lambda}(t)$ be continuously time-dependent, and consider the time-*dependent* Schrödinger equation

$$i\hbar \frac{d}{dt}\,|\,\Psi(t)\,\rangle = H\big(\boldsymbol{\lambda}(t)\big)\,|\,\Psi(t)\,\rangle \quad . \tag{1.174}$$

The adiabatic theorem states that if $\boldsymbol{\lambda}(t)$ evolves extremely slowly, then each solution $|\,\Psi_n(t)\,\rangle$ is proportional to $|\,\varphi_n\big(\boldsymbol{\lambda}(t)\big)\,\rangle$, with

$$|\,\Psi_n\big(\boldsymbol{\lambda}(t)\big)\,\rangle = \exp\big(i\gamma_n(t)\big)\,\exp\Big(-\frac{i}{\hbar}\int^t dt'\,E_n\big(\boldsymbol{\lambda}(t')\big)\Big)\,|\,\varphi_n\big(\boldsymbol{\lambda}(t)\big)\,\rangle \quad , \tag{1.175}$$

with corrections which vanish in the limit $|\dot{\boldsymbol{\lambda}}|/|\boldsymbol{\lambda}| \to 0$. Taking the time derivative and then the overlap with the bra vector $\langle\, \varphi_n\big(\boldsymbol{\lambda}(t)\big)\,|$, one obtains the result

$$\frac{d\gamma_n(t)}{dt} = i\,\langle\, \varphi_n\big(\boldsymbol{\lambda}(t)\big)\,|\,\frac{d}{dt}\,|\,\varphi_n\big(\boldsymbol{\lambda}(t)\big)\,\rangle = \boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda})\cdot\frac{d\boldsymbol{\lambda}}{dt} \equiv \mathcal{A}_n(t) \quad , \tag{1.176}$$

where

$$\mathcal{A}_n^\mu(\boldsymbol{\lambda}) = i\,\langle\, \varphi_n(\boldsymbol{\lambda})\,|\,\frac{\partial}{\partial\lambda_\mu}\,|\,\varphi_n(\boldsymbol{\lambda})\,\rangle \quad . \tag{1.177}$$

Note that $\mathcal{A}_n^\mu(\boldsymbol{\lambda})$ is real. In particular, if $\boldsymbol{\lambda}(t)$ traverses a closed loop $\mathcal{C}$ with infinitesimal speed, then the wavefunction $|\,\Psi_n(t)\,\rangle$ will accrue a *geometric phase* $\gamma_n(\mathcal{C})$, given by

$$\gamma_n(\mathcal{C}) = \oint\limits_{\mathcal{C}} d\boldsymbol{\lambda}\cdot\boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda}) \quad , \tag{1.178}$$

called *Berry's phase*[28].

In the adiabatic limit, the dynamical phase $\hbar^{-1}\int^t dt'\; E_n\big(\boldsymbol{\lambda}(t')\big)$ becomes very large if $E_n \neq 0$, because the path $\boldsymbol{\lambda}(t)$ is traversed very slowly. We may remove this dynamical phase by defining the Hamiltonian

$$\widetilde{H}_n(\boldsymbol{\lambda}) \equiv H(\boldsymbol{\lambda}) - E_n(\boldsymbol{\lambda}) \quad . \tag{1.179}$$

We define $|\,\widetilde{\Psi}_n(t)\,\rangle$ as the solution to the Schrödinger equation

$$i\hbar\,\frac{d}{dt}\,|\,\widetilde{\Psi}_n(t)\,\rangle = \widetilde{H}_n\big(\boldsymbol{\lambda}(t)\big)\,|\,\widetilde{\Psi}_n(t)\,\rangle \tag{1.180}$$

in the adiabatic limit. The adiabatic wavefunctions $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle$ are the same as before, but now satisfy the zero energy condition $\widetilde{H}_n(\boldsymbol{\lambda})\,|\,\varphi_n(\boldsymbol{\lambda})\,\rangle = 0$. Clearly $|\,\widetilde{\Psi}_n\big(\boldsymbol{\lambda}(t)\big)\,\rangle = \exp\big(i\gamma_n(t)\big)\,|\,\varphi_n\big(\boldsymbol{\lambda}(t)\big)\,\rangle$ and the dynamical phase has been removed. However, note that the geometrical phase $\gamma_n$ does not depend on the elapsed time, but only on the path traversed, *viz.*

$$\gamma_n = \gamma_n(\boldsymbol{\lambda}) = \int\limits_{\boldsymbol{\lambda}_0}^{\boldsymbol{\lambda}} d\boldsymbol{\lambda}'\cdot\boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda}') \quad , \tag{1.181}$$

where $\boldsymbol{\lambda}_0 = \boldsymbol{\lambda}(0)$, and where the integral is taken along the path in traversed by $\boldsymbol{\lambda}$.

### 1.7.2 Connection and curvature

The mathematical structure underlying this discussion is that of the *Hermitian line bundle*, the ingredients of which are (i) a *base space* $\mathcal{M}$ which is a topological manifold; this is the parameter

[28] See M. V. Berry, *Proc. Roy. Soc. A* **392**, 45 (1984).

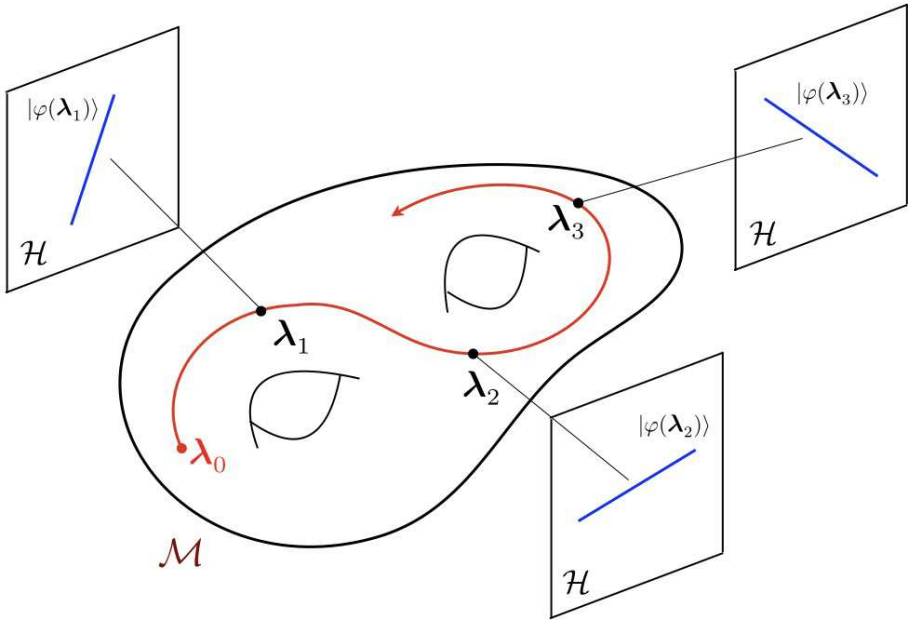


Figure 1.20: A Hermitian line bundle consists of a base space $\mathcal{M}$ and a fiber $|\,\varphi(\boldsymbol{\lambda})\,\rangle$ which twists as the point $\boldsymbol{\lambda}$ moves around the base space.

space for $\boldsymbol{\lambda}$, and (ii) to each point $\boldsymbol{\lambda} \in \mathcal{M}$ is associated a *fiber* which is the adiabatic wavefunction $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle \in \mathcal{H}$, which is a complex one-dimensional subspace of some Hilbert space $\mathcal{H}$. As $\boldsymbol{\lambda}$ moves around the base space $\mathcal{M}$, the fiber twists around. The adiabatic theorem furnishes us with a way of defining *parallel transport* of $|\,\widetilde{\Psi}_n(\boldsymbol{\lambda})\,\rangle$ along the curve $\mathcal{C}$[29]. The object $\boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda})$ is the *connection* and the geometric phase $\gamma(\mathcal{C})$ is the *holonomy* of the connection[30]. As a holonomy, $\gamma(\mathcal{C})$ depends only on the curve $\mathcal{C}$ and not on where along the curve one starts.

The *curvature tensor* for the bundle is given by

$$\begin{aligned}\Omega_n^{\mu\nu}(\boldsymbol{\lambda}) &= \frac{\partial \mathcal{A}_n^\nu}{\partial \lambda_\mu} - \frac{\partial \mathcal{A}_n^\mu}{\partial \lambda_\nu} \\ &= i\left\langle \frac{\partial \varphi_n}{\partial \lambda_\mu} \,\middle|\, \frac{\partial \varphi_n}{\partial \lambda_\nu} \right\rangle - i\left\langle \frac{\partial \varphi_n}{\partial \lambda_\nu} \,\middle|\, \frac{\partial \varphi_n}{\partial \lambda_\mu} \right\rangle \quad .\end{aligned} \tag{1.182}$$

Using completeness of the $|\,\varphi_n\,\rangle$ basis, we may write the curvature tensor as

$$\Omega_n^{\mu\nu}(\boldsymbol{\lambda}) = i \sum_l{}' \frac{\left\langle \varphi_n \,\middle|\, \frac{\partial H}{\partial \lambda_\mu} \,\middle|\, \varphi_l \right\rangle \left\langle \varphi_l \,\middle|\, \frac{\partial H}{\partial \lambda_\nu} \,\middle|\, \varphi_n \right\rangle - (\mu \leftrightarrow \nu)}{(E_n - E_l)^2} \quad , \tag{1.183}$$

[29]Note that $|\,\Psi_n(t)\,\rangle$, which depends explicitly on elapsed time and not solely on the position $\boldsymbol{\lambda}$ along its trajectory, can not be said to be parallel transported along any curve.

[30]See B. Simon, *Phys. Rev. Lett.* **51**, 2167 (1983).

where the prime on the sum indicates that the term $l = n$ is to be excluded. We see that in this formulation the curvature tensor is actually independent of any phase convention for the adiabatic wavefunctions $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle$. So long as the adiabatic eigenstate $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle$ remains nondegenerate, the denominator in Eqn. 1.183 remains nonzero, hence the curvature tensor $\Omega(\boldsymbol{\lambda})$ is nonsingular. The same cannot be said about the connection $\boldsymbol{\mathcal{A}}(\boldsymbol{\lambda})$, however, because it is *gauge-covariant*. This means that if we multiply the adiabatic wavefunctions by phases, with $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle \to \exp(if_n(\boldsymbol{\lambda}))\,|\,\varphi_n(\boldsymbol{\lambda})\,\rangle$, the connection changes accordingly, with

$$\boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda}) \to \boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda}) - \frac{\partial f_n(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} \quad . \tag{1.184}$$

How can we fix a gauge in order to give unambiguous meaning to $\boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda})$? One way might be to demand that the adiabatic wavefunction amplitude be real and positive at some particular point in space $\boldsymbol{r}_0$, *i.e.* $\langle\,\boldsymbol{r}_0\,|\,\varphi_n(\boldsymbol{\lambda})\,\rangle \in \mathbb{R}_+$ for all $\boldsymbol{\lambda} \in \mathcal{M}$. For lattice-based models, where the adiabatic wavefunction is a vector of amplitudes for each orbital and each site within the appropriate unit cell, we could similarly demand that one of these amplitudes be real and positive. *This prescription fails if there exists a value of $\boldsymbol{\lambda}$ for which this wavefunction amplitude vanishes.*

As we are about to discover, the integral of the curvature over a two-dimensional base space is a topological invariant, meaning that it remains fixed (and indeed quantized) under continuous deformations of the Hamiltonian $H(\boldsymbol{\lambda})$. Using Stokes' theorem, we can turn an area integral of the curvature into line integrals of the connection. However, having chosen a particular gauge for the adiabatic wavefunctions, it may be that the connection is singular at certain points. Therefore the line integrals cannot be completely collapsed, and we obtain the result

$$\int\limits_{\mathcal{M}} d^2\lambda\; \Omega_n^{12}(\boldsymbol{\lambda}) = -\sum_i \oint\limits_{\mathcal{C}_i} d\boldsymbol{\lambda} \cdot \boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda}) \quad , \tag{1.185}$$

where the loop $\mathcal{C}_i$ encloses the $i^{\rm th}$ singularity $\boldsymbol{\lambda}_i$ of the connection in a counterclockwise manner[31]. This is the generalization to Hermitian line bundles of the index formula in Eqn. 1.288 for the Gauss-Bonnet theorem. Quantization follows by writing $|\,\varphi_n(\boldsymbol{\lambda})\,\rangle = e^{iq_i\zeta(\boldsymbol{\lambda}-\boldsymbol{\lambda}_i)}\,|\,\widetilde{\varphi}_n(\boldsymbol{\lambda})\,\rangle$ in the vicinity of $\boldsymbol{\lambda} = \boldsymbol{\lambda}_i$, where $q_i$ is an integer and

$$\zeta(\boldsymbol{\lambda} - \boldsymbol{\lambda}_i) = \tan^{-1}\left(\frac{\lambda_2 - \lambda_{i,2}}{\lambda_1 - \lambda_{i,1}}\right) \quad . \tag{1.186}$$

The integers $q_i$ are chosen to 'unwind' the singularity at each $\boldsymbol{\lambda}_i$, so as to make the gauge transformed connection $\widetilde{\boldsymbol{\mathcal{A}}}_n^{\mu}(\boldsymbol{\lambda}) \equiv i\,\langle\,\widetilde{\varphi}_n(\boldsymbol{\lambda})\,|\,\boldsymbol{\nabla}_{\boldsymbol{\lambda}}\,|\,\widetilde{\varphi}_n(\boldsymbol{\lambda})\,\rangle$ nonsingular[32]. We then obtain

$$C_n \equiv \frac{1}{2\pi}\int\limits_{\mathcal{M}} d^2\lambda\; \Omega_n^{12}(\boldsymbol{\lambda}) = \sum_i q_i \quad . \tag{1.187}$$

[31] We must assume that the base space $\mathcal{M}$ is orientable.

[32] Note that we have employed a singular gauge transformation, which is necessary to do the desired unwinding. Also note that the integers $q_i$ should also carry a band index $n$, which we have suppressed here for notational simplicity.

Thus $C_n \in \mathbb{Z}$ is the *Chern number* of the Hermitian line bundle corresponding to the adiabatic wavefunction $|\,\varphi_n\,\rangle$.

The simplest nontrivial example is that of a spin-$\frac{1}{2}$ object in a magnetic field $\boldsymbol{B}(t)$, with

$$H(t) = g\mu_{\rm B}\boldsymbol{B}\cdot\boldsymbol{\sigma} = g\mu_{\rm B}B\begin{pmatrix}\cos\theta & \sin\theta\exp(-i\phi)\\ \sin\theta\exp(i\phi) & \cos\theta\end{pmatrix} \quad , \tag{1.188}$$

where $\boldsymbol{B} = B\,\hat{\boldsymbol{n}}$ is the adiabatic parameter which varies extremely slowly in time. The adiabatic wavefunctions are

$$|\,\varphi_+(\theta,\phi)\,\rangle = \begin{pmatrix}u\\ v\end{pmatrix} \qquad , \qquad |\,\varphi_-(\theta,\phi)\,\rangle = \begin{pmatrix}-\bar{v}\\ \bar{u}\end{pmatrix} \quad , \tag{1.189}$$

where $\hat{\boldsymbol{n}} = (\sin\theta\cos\phi\,,\,\sin\theta\sin\phi\,,\,\cos\theta)$, $u = \cos(\frac{1}{2}\theta)$, and $v = \sin(\frac{1}{2}\theta)\exp(i\phi)$. The energy eigenvalues are $E_\pm = \pm g\mu_{\rm B}B$. We now compute the connections,

$$\begin{aligned}\mathcal{A}_+ &= i\,\langle\,\varphi_+\,|\,\frac{d}{dt}\,|\,\varphi_+\,\rangle = i\big(\bar{u}\dot{u} + \bar{v}\dot{v}\big) = -\tfrac{1}{2}(1-\cos\theta)\,\dot{\phi} = -\tfrac{1}{2}\,\dot{\omega}\\ \mathcal{A}_- &= i\,\langle\,\varphi_-\,|\,\frac{d}{dt}\,|\,\varphi_-\,\rangle = i\big(u\dot{\bar{u}} + v\dot{\bar{v}}\big) = +\tfrac{1}{2}(1-\cos\theta)\,\dot{\phi} = +\tfrac{1}{2}\,\dot{\omega} \quad ,\end{aligned} \tag{1.190}$$

where $\dot{\omega}$ is the differential solid angle subtended by the path $\hat{\boldsymbol{n}}(t)$. Thus, $\gamma_\pm(\mathcal{C}) = \mp\frac{1}{2}\,\omega_{\mathcal{C}}$ is $\mp$ half the solid angle subtended by the path $\hat{\boldsymbol{n}}_{\mathcal{C}}(t)$ on the Bloch sphere. We may now read off the components $\mathcal{A}^\theta_\pm = 0$ and $\mathcal{A}^\phi_\pm = \mp\frac{1}{2}(1-\cos\theta)$ and invoke Eqn. 1.182 to compute the curvature,

$$\Omega^{\theta\phi}_\pm(\theta,\phi) = \mp\tfrac{1}{2}\sin\theta \quad . \tag{1.191}$$

Note then that the integral of the curvature over the entire sphere is given by

$$\int\limits_0^{2\pi}\!d\phi\int\limits_0^{\pi}\!d\theta\;\Omega^{\theta\phi}_\pm(\theta,\phi) = 2\pi\,C_\pm \quad , \tag{1.192}$$

where $C_\pm = \mp 1$ is the Chern number. Equivalently, note that both connections are singular at $\theta = \pi$, where the azimuthal angle is ill-defined. This singularity can be gentled through an appropriate singular gauge transformation $|\,\varphi_\pm\,\rangle = e^{\pm i\phi}\,|\,\widetilde{\varphi}_\pm\,\rangle = e^{\mp i\zeta}\,|\,\widetilde{\varphi}_\pm\,\rangle$, where $\zeta$ is defined to be the angle which increases as one winds counterclockwise around the *south pole*, hence $\zeta = -\phi$. This corresponds to $q_\pm = \mp 1$ in our earlier notation, hence again $C_\pm = \mp 1$.

As we saw above, this is a general result: when the base space $\mathcal{M}$ is two dimensional: the integral of the curvature over $\mathcal{M}$ is $2\pi$ times an integer. This result calls to memory the famous Gauss-Bonnet theorem (see §1.10 below for more), which says that the integral of the Gaussian curvature $K$ over a two-dimensional manifold $\mathcal{M}$ is

$$\int\limits_{\mathcal{M}} dS\;K = 4\pi(1-g) \quad , \tag{1.193}$$

where $g$ is the *genus* (number of holes) in the manifold $\mathcal{M}$. In the Gauss-Bonnet case, the bundle construction is known as the *tangent bundle* of $\mathcal{M}$, and the corresponding connection is determined by the Riemannian metric one places on $\mathcal{M}$. However, *independent of the metric*, the integral of $K$ is determined solely by the global topology of $\mathcal{M}$, *i.e.* by its genus. Thus, in three-dimensional space, a sphere inherits a metric from its embedding in $\mathbb{R}^3$. If you distort the sphere by denting it, locally its curvature $K$ will change, being the product of the principal radii of curvature at any given point. But the integral of $K$ over the surface will remain fixed at $4\pi$. Just as the genus $g$ of a Riemann surface is unaffected by simple deformations but can change if one does violence to it, such as puncturing and resewing it[33], so is the Chern number invariant under deformations of the underlying Hamiltonian, provided one does not induce a level crossing of the adiabatic eigenstate $|\,\varphi_n\,\rangle$ with a neighboring one. Also, note that if the connection $\boldsymbol{\mathcal{A}}_n(\boldsymbol{\lambda})$ can be defined globally on $\mathcal{M}$, *i.e.* with no singularities, then $C_n = 0$.

### 1.7.3 Two-band models

For the two band ($S = \frac{1}{2}$) system with Hamiltonian $H = g\mu_{\rm B} B\,\hat{\boldsymbol{n}}(\boldsymbol{\lambda})\cdot\boldsymbol{\sigma}$, one can verify that we may also write the Chern numbers as[34]

$$C_\pm = \pm\frac{1}{4\pi}\int\limits_{\mathcal{M}} d^2\lambda\;\hat{\boldsymbol{n}}\cdot\frac{\partial\hat{\boldsymbol{n}}}{\partial\lambda_1}\times\frac{\partial\hat{\boldsymbol{n}}}{\partial\lambda_2}\quad . \tag{1.194}$$

In this formulation, the Chern number has the interpretation of a *Pontrjagin number*, which is a topological index classifying real vector bundles (more in §1.10.2 below). Thus, for a tight binding model on a bipartite lattice, the most general Hamiltonian may be written

$$H(\boldsymbol{k}) = d_0(\boldsymbol{k}) + \boldsymbol{d}(\boldsymbol{k})\cdot\boldsymbol{\sigma}\quad , \tag{1.195}$$

where $\boldsymbol{k}$ is the wavevector and where each $d^\mu(\boldsymbol{k})$ is periodic under translations of $\boldsymbol{k}$ by any reciprocal lattice vector $\boldsymbol{G}$. In this case $\lambda_{1,2} = k_{x,y}$ are the components of $\boldsymbol{k}$, and $\mathcal{M} = \mathbb{T}^2$ is the Brillouin zone torus. Note that the sum of the Chern numbers for each of the $+$ and $-$ bands is zero. As we shall see below with the TKNN problem, for a larger spin generalization, *i.e.* when the magnetic unit cell contains more than two basis elements, the sum $\sum_a C_a$ of the Chern numbers over all bands also vanishes. Consider the two band model with

$$H(\boldsymbol{\theta}) = \begin{pmatrix} m - 2t\cos\theta_1 - 2t\cos\theta_2 & \Delta\,(\sin\theta_1 - i\sin\theta_2) \\ \Delta\,(\sin\theta_1 + i\sin\theta_2) & -m + 2t\cos\theta_1 + 2t\cos\theta_2 \end{pmatrix}\quad . \tag{1.196}$$

As before, $\theta_\mu = \boldsymbol{k}\cdot\boldsymbol{a}_\mu$. Note $H(\boldsymbol{\theta}) = \boldsymbol{d}(\boldsymbol{\theta})\cdot\boldsymbol{\sigma}$ with

$$\begin{aligned}\boldsymbol{d}(\boldsymbol{\theta}) &= \big(\Delta\sin\theta_1\,,\,\Delta\sin\theta_2\,,\,m - 2t\cos\theta_1 - 2t\cos\theta_2\big) \\ &\equiv |\boldsymbol{d}|\,\big(\sin\vartheta\cos\phi\,,\,\sin\vartheta\sin\phi\,,\,\cos\vartheta\big)\quad .\end{aligned} \tag{1.197}$$

[33] M. Gilbert's two commandments of topology: (I) Thou shalt not cut. (II) Thou shalt not glue.

[34] The dependence of the magnitude $B = |\boldsymbol{B}|$ on $\boldsymbol{\lambda}$ is irrelevant to the calculation of the Chern numbers. The equivalence $\hat{\boldsymbol{n}} = z^\dagger\boldsymbol{\sigma}z$ for $z = \begin{pmatrix} u \\ v\end{pmatrix}$ is known as the *first Hopf map* from $\mathsf{CP}^1$ to $\mathbb{S}^2$.

Note the adiabatic parameters here are $\theta_1$ and $\theta_2$, upon which $\vartheta$ and $\phi$ are parametrically dependent. Does $\boldsymbol{d}(\boldsymbol{\theta})$ wind around the Brillouin zone torus, yielding a nonzero Chern number?

First, you might be wondering, where does this model come from? Actually, it is the Hamiltonian for a $p_x + ip_y$ superconductor, but we can back out of $H(\boldsymbol{\theta})$ a square lattice insulator model involving two orbitals $a$ and $b$ which live on top of each other at each site, and are not spatially separated[35]. The parameter $m$ reflects the difference in the local energies of the two orbitals. The nearest neighbor hopping integrals between like orbitals are $t_{aa} = t$ and $t_{bb} = -t$, but $t_{ab}(\pm\boldsymbol{a}_1) = \pm\frac{i}{2}\,\Delta$ and $t_{ab}(\pm\boldsymbol{a}_2) = \pm\frac{1}{2}\Delta$, with $t_{ba}(-\boldsymbol{a}_{1,2}) = t^*_{ab}(+\boldsymbol{a}_{1,2})$ due to hermiticity.

The energy eigenvalues are

$$E_\pm(\boldsymbol{\theta}) = \pm\sqrt{\Delta^2\sin^2\theta_1 + \Delta^2\sin^2\theta_2 + (m - 2t\cos\theta_1 - 2t\cos\theta_2)^2} \quad . \tag{1.198}$$

The Wigner-von Neumann theorem says that degeneracy for complex Hamiltonians like ours has codimension three, meaning one must fine tune three parameters in order to get a degeneracy. The reason is that for $H = \boldsymbol{d}\cdot\sigma$ describing two nearby levels, the gap is $2|\boldsymbol{d}|$, thus in order for the gap to vanish we must require three conditions: $d_x = d_y = d_z = 0$. For the real case where $d_y = 0$ is fixed, we only require two conditions, *i.e.* $d_x = d_z = 0$. For our model, the gap collapse requires

$$\begin{aligned} \Delta\,\sin\theta_1 &= 0 \\ \Delta\,\sin\theta_2 &= 0 \\ m - 2t\cos\theta_1 - 2t\cos\theta_2 &= 0 \quad . \end{aligned} \tag{1.199}$$

Thus, degeneracies occur at $(\theta_1,\theta_2) = (0,0)$ when $m = 4t$, at $(\pi,\pi)$ when $m = -4t$, and at $(0,\pi)$ and $(\pi,0)$ when $m = 0$. It is clear that for $|m| > 4t$ both Chern numbers must be zero. This is because for $m > 4t$ we have $d_z(\theta_1,\theta_2) > 0$ for all values of the Bloch phases, while for $m < -4t$ we have $d_z(\theta_1,\theta_2) < 0$. Thus in neither case can the $\boldsymbol{d}$ vector wind around the Bloch sphere, and the Pontrjagin/Chern indices accordingly vanish for both bands.

Now consider the case $m \in [0, 4t]$. Recall that the eigenfunctions are given by

$$|\,\varphi_+\,\rangle = \begin{pmatrix} \cos(\frac{1}{2}\vartheta) \\ \sin(\frac{1}{2}\vartheta)\,e^{i\chi} \end{pmatrix} \qquad , \qquad |\,\varphi_-\,\rangle = \begin{pmatrix} -\sin(\frac{1}{2}\vartheta)\,e^{-i\chi} \\ \cos(\frac{1}{2}\vartheta) \end{pmatrix} \quad , \tag{1.200}$$

with eigenvalues $\pm|\boldsymbol{d}|$. The singularity in both $|\,\varphi_\pm(\theta_1,\theta_2)\,\rangle$ occurs at $\vartheta = \pi$. Recall that $\boldsymbol{d} \equiv |\boldsymbol{d}|\big(\sin\vartheta\cos\chi\,,\,\sin\vartheta\sin\chi\,,\,\cos\vartheta\big)$, which entails $\boldsymbol{d} = (0\,,\,0\,,\,-|\boldsymbol{d}|)$, *i.e.* $d_x = d_y = 0$ and $d_z < 0$. This only occurs for $(\theta_1,\theta_2) = (0,0)$. All we need to do to compute the Chern numbers is to identify the singularity in $\zeta(\theta_1,\theta_2)$ about this point, *i.e.* does $\zeta = -\chi$ wind clockwise or counterclockwise, in which case $C_+ = -1$ or $C_+ = +1$, respectively. Treating $\theta_{1,2}$ as very small, one easily obtains $\zeta = -\tan^{-1}(\theta_2/\theta_1)$, which is to say clockwise winding, hence $C_\pm = \mp 1$. *Exercise*: Find $C_\pm$ for $m \in [-4t, 0]$.

[35] In this model they are both $s$-orbitals, which is unphysical.

### Haldane honeycomb model

In §1.6.3, we met Haldane's famous honeycomb lattice model, $H(\boldsymbol{\theta}) = d_0(\boldsymbol{\theta}) + \boldsymbol{d}(\boldsymbol{\theta}) \cdot \boldsymbol{\sigma}$, with

$$\begin{aligned} d_0(\boldsymbol{\theta}) &= -2t_2\big[\cos\theta_1 + \cos\theta_2 + \cos(\theta_1+\theta_2)\big]\cos\phi \\ d_x(\boldsymbol{\theta}) &= -t_1\big(1 + \cos\theta_1 + \cos\theta_2\big) \\ d_y(\boldsymbol{\theta}) &= t_1\big(\sin\theta_1 - \sin\theta_2\big) \\ d_z(\boldsymbol{\theta}) &= m - 2t_2\big[\sin\theta_1 + \sin\theta_2 - \sin(\theta_1+\theta_2)\big]\sin\phi \quad . \end{aligned} \tag{1.201}$$

The energy eigenvalues are $E_\pm(\boldsymbol{\theta}) = d_0(\boldsymbol{\theta}) \pm |\boldsymbol{d}(\boldsymbol{\theta})|$. Now is quite easy to demonstrate that $\big|\sin\theta_1 + \sin\theta_2 - \sin(\theta_1+\theta_2)\big| \le \frac{3}{2}\sqrt{3}$, and therefore that the $\boldsymbol{d}(\boldsymbol{\theta})$ cannot wind if $|m| > 3\sqrt{3}\,t_2\,|\sin\phi|$ and $C_\pm = 0$. As above, we set $\boldsymbol{d} \equiv |\boldsymbol{d}|\big(\sin\vartheta\cos\chi\,,\,\sin\vartheta\sin\chi\,,\,\cos\vartheta\big)$, and the singularity in both wavefunctions occurs at $\vartheta = \pi$, which requires $d_x(\boldsymbol{\theta}) = d_y(\boldsymbol{\theta}) = 0$ and $d_z(\boldsymbol{\theta}) < 0$. This in turn requires $\theta_1 = \theta_2 = \frac{2}{3}\pi s$ where $s = \mathsf{sgn}(\sin\phi)$. We now write $\theta_j = \frac{2}{3}\,\pi s + \delta_j$ and find

$$\tan\chi = \frac{d_y(\boldsymbol{\theta})}{d_x(\boldsymbol{\theta})} = \frac{\delta_2 - \delta_1}{\delta_1 + \delta_2}\,\mathsf{sgn}(\sin\phi) = \tan(\alpha - \tfrac{\pi}{4})\,\mathsf{sgn}(\sin\phi) \quad , \tag{1.202}$$

where $\boldsymbol{\delta} \equiv |\boldsymbol{\delta}|\,(\cos\alpha, \sin\alpha)$. Thus the winding of $\zeta = -\chi$ is in the sense of that of $\alpha$ if $s < 0$ and opposite if $s > 0$, and we conclude $C_\pm = \mp\mathsf{sgn}(\sin\phi)$. The topological phase diagram for the Haldane honeycomb lattice model is shown in Fig. 1.21. The phase space is a cylinder in the dimensionless parameters $\phi \in [-\pi, \pi]$ and $m/t_2 \in \mathbb{R}$. Regions are labeled by the Chern numbers $C_\pm$ of the two energy bands.

### Note on broken symmetries

In §1.4.3 we derived the long wavelength Hamiltonian for the $\boldsymbol{K}$ and $\boldsymbol{K}' = -\boldsymbol{K}$ valleys of graphene. If we adopt a pseudospin convention for the valleys, with Pauli matrices $\boldsymbol{\tau}$, and where $\tau^z = \pm 1$ corresponds to the $\pm\boldsymbol{K}$ valley, we may in one stroke write the long wavelength graphene Hamiltonian as

$$H_0 = \tfrac{\sqrt{3}}{2}\,ta\big(q_x\,\sigma^x\,\tau^z + q_y\,\sigma^y\big) \quad . \tag{1.203}$$

This Hamiltonian is symmetric under the operations of parity ($\mathcal{P}$) and time-reversal ($\mathcal{T}$). Under $\mathcal{P}$, we switch valleys, switch sublattices, and send $q_x \to -q_x$. Under $\mathcal{T}$, we switch valleys and send $\boldsymbol{q} \to -\boldsymbol{q}$. It is also important to remember that $\mathcal{T}$ is antiunitary, and includes the complex conjugation operator $\check{K}$. The matrix parts of these operators, *i.e.* other than their actions on the components of $\boldsymbol{q}$, are given by

$$\mathcal{P} = \sigma^y\,\tau^z \qquad , \qquad \mathcal{T} = i\tau^y\check{K} \quad . \tag{1.204}$$

Note that $\mathcal{T}^2 = -1$ and $\mathcal{T}^{-1} = -\mathcal{T} = \check{K}\tau^y\,(-i)$. Of course $\mathcal{P}^2 = 1$ and thus $\mathcal{P}^{-1} = \mathcal{P}$. One can now check explicitly that $\mathcal{P}H_0\mathcal{P}^{-1} = \mathcal{T}H_0\mathcal{T}^{-1} = H_0$.

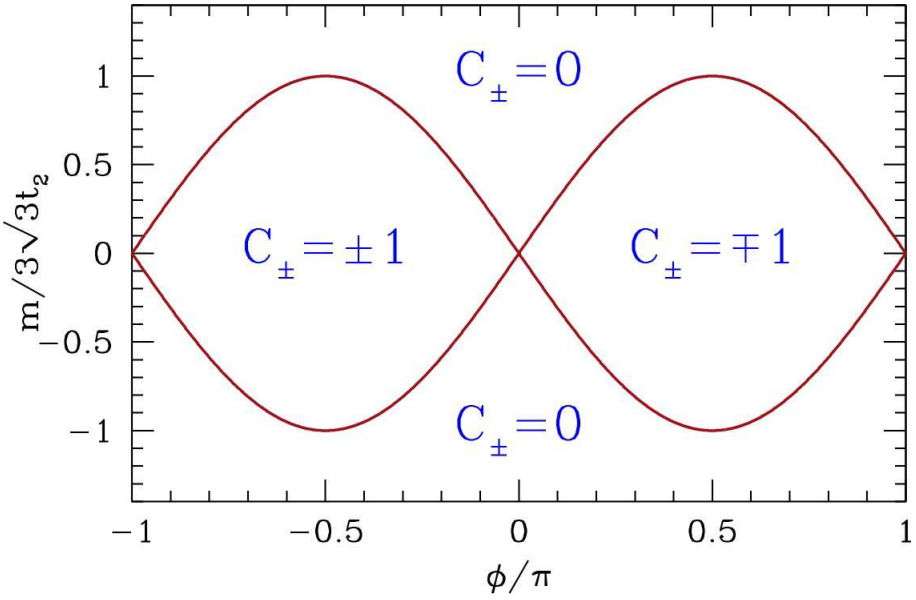


Figure 1.21: Topological phase diagram for the Haldane honeycomb lattice model, in which regions in the $(\sin\phi, m/t_2)$ cylinder are labeled by the Chern numbers $C_\pm$ of the energy bands.

There are three ways to introduce a gap into the model, *i.e.* to gap out the dispersion at the $\boldsymbol{K}$ and $\boldsymbol{K}'$ points at the two inequivalent Brillouin zone corners:

1. The first way is by introducing a Semenoff mass term, which is of the form $V_{\rm S} = \Delta_{\rm S}\,\sigma^z$. This turns graphene into boron nitride (BN), distinguishing the local $\pi$-orbital energies on the B and N sites. One can check that

$$\mathcal{P}\,\sigma^z\,\mathcal{P}^{-1} = -\sigma^z \quad , \qquad \mathcal{T}\,\sigma^z\,\mathcal{T}^{-1} = +\sigma^z \quad , \tag{1.205}$$

   and therefore the Semenoff mass breaks parity and preserves time-reversal.

2. The second way comes from the Haldane honeycomb lattice model at $m = 0$, where $V_{\rm H} = \Delta_{\rm H}\,\sigma^z\tau^z$.

$$\mathcal{P}\,\sigma^z\tau^z\,\mathcal{P}^{-1} = +\sigma^z\tau^z \quad , \qquad \mathcal{T}\,\sigma^z\tau^z\,\mathcal{T}^{-1} = -\sigma^z\tau^z \quad . \tag{1.206}$$

   This term, the Haldane mass, preserves parity but breaks time-reversal. It leads to a topological band structure in which the bands are classified by nonzero Chern numbers.

3. The third way involves introducing the physical electron spin, and arises from spin-orbit effects. It essentially is described by two copies of the Haldane model, in which the up and down spin electrons couple oppositely to magnetic flux. This was first discussed by Kane and Mele[36], and is described by the perturbation $V_{\rm KM} = \Delta_{\rm KM}\,\sigma^z\tau^z s^z$. where $\boldsymbol{s}$ is the

[36]C. L. Kane and E. J. Mele, *Phys. Rev. Lett.* **95**, 226801 (2005).

electron spin operator . The Kane-Mele mass term preserves $\mathcal{P}$ and $\mathcal{T}$ symmetries:

$$\mathcal{P}\,\sigma^z\tau^z s^z\,\mathcal{P}^{-1} = +\sigma^z\tau^z s^z \quad , \quad \mathcal{T}\,\sigma^z\tau^z s^z\,\mathcal{T}^{-1} = +\sigma^z\tau^z s^z \quad . \tag{1.207}$$

Therefore, following the tried and true rule in physics that "everything which is not forbidden is compulsory", there *must* be a KM mass term in real graphene. The catch is that it is extremely small because graphene is a low-$Z$ atom, and first principles calculations[37] conclude that the spin-orbit gap is on the order of $10\,\mathrm{mK}$ – too small to be observed due to finite temperature and disorder effects. However, there are many materials (Bi bilayers, HgTe/CdTe heterostructures, various three-dimensional materials such as $\alpha$-Sn, $\mathrm{Bi}_x\mathrm{Sb}_{1-x}$ and others) where the effect is predicted to be sizable and where it is indeed observed. This is the essence of topological insulator behavior.

### 1.7.4 The TKNN formula

Recall the Hamiltonian of Eqn. 1.163 for the isotropic square lattice Hofstadter model with flux $\phi = 2\pi p/q$ per unit cell. A more general version, incorporating anisotropy which breaks $90^\circ$ rotational symmetry, is given by[38]

$$H(\theta_1,\theta_2) = -\begin{pmatrix} 2t_2\cos\theta_2 & t_1 & 0 & \cdots & 0 & t_1\,e^{-i\theta_1} \\ t_1 & 2t_2\cos\left(\theta_2+\frac{2\pi p}{q}\right) & t_1 & & & 0 \\ 0 & t_1 & 2t_2\cos\left(\theta_2+\frac{4\pi p}{q}\right) & t_1 & & \vdots \\ \vdots & 0 & t_1 & \ddots & & \vdots \\ & & & & & t_1 \\ 0 & & & \cdots & t_1 & 2t_2\cos\left(\theta_2+\frac{2\pi(q-1)p}{q}\right) \\ t_1\,e^{i\theta_1} & 0 & & & & \end{pmatrix} . \tag{1.208}$$

This is a $q\times q$ matrix, and the $q$ eigenvectors $|\,\varphi_n(\boldsymbol{\theta})\,\rangle$ are labeled by a band index $n\in\{1,\ldots,q\}$, with component amplitudes $\varphi_{a,n}(\boldsymbol{\theta})$ satisfying

$$H_{aa'}(\boldsymbol{\theta})\,\varphi_{a',n}(\boldsymbol{\theta}) = E_n(\boldsymbol{\theta})\,\varphi_{a,n}(\boldsymbol{\theta}) \quad . \tag{1.209}$$

From Wigner-von Neumann, we expect generically that neighboring bands will not cross as a function of the two parameters $(\theta_1,\theta_2)$, because degeneracy has codimension three. Thus, associated with each band $n$ is a Chern number $C_n$. By color coding each spectral gap according to the Chern number of all bands below it, J. Avron produced a beautiful and illustrative image of Hofstadter's butterfly, shown in Fig. 1.22 for the isotropic square lattice and in Fig. 1.23 for the isotropic honeycomb lattice.

[37] See Y. Yao *et al.*, *Phys. Rev. B* **75**, 041401(R) (2007).

[38] We drop the hat on $\hat{H}(\boldsymbol{\theta})$ but fondly recall that $\hat{H}_{aa'}(\boldsymbol{\theta})$ is the lattice Fourier transform of $H_{aa'}(\boldsymbol{R}-\boldsymbol{R}')$.

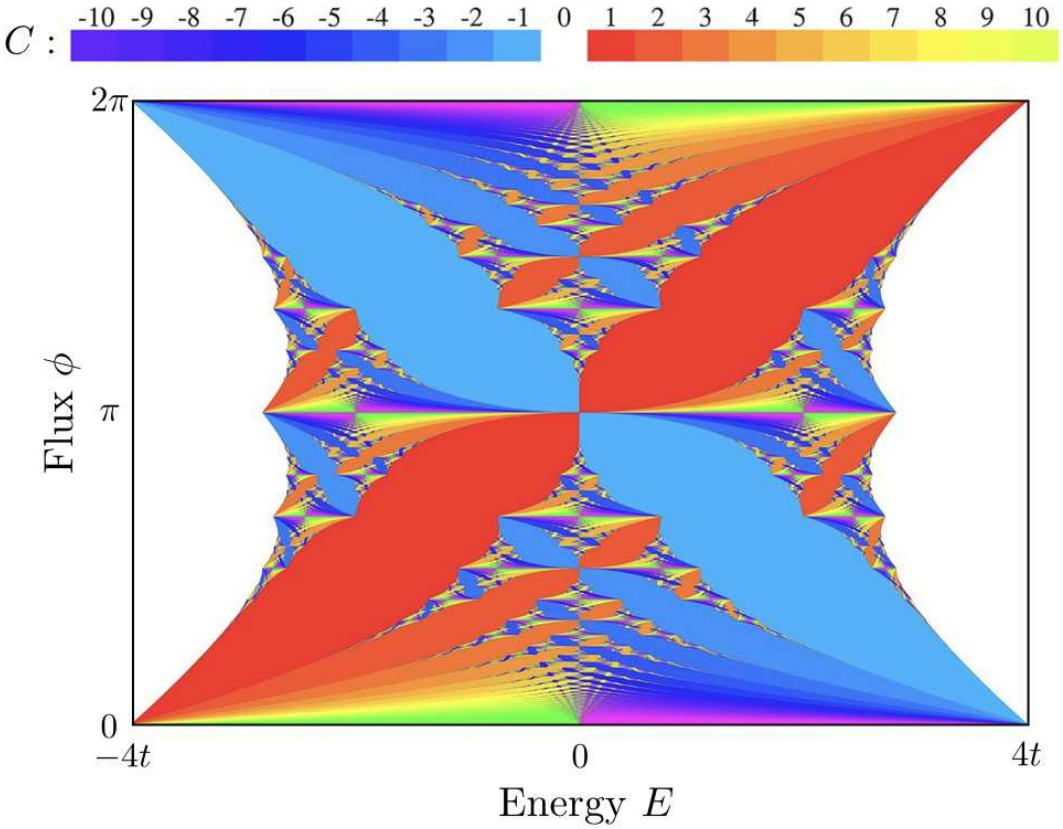


Figure 1.22: Avron's depiction of the Hofstadter butterfly for the isotropic square lattice system. The band gap regions are color coded by Chern number, $C$, which is the sum of the Chern numbers of all bands below a given gap. White regions correspond to $C = 0$. See J. E. Avron, *Colored Hofstadter butterflies*, in *Multiscale Methods in Quantum Mechanics*, P. Blanchard and G. Dell'Antonio, eds. (Birkhäuser, 2004).

It turns out that the Chern number is not just an abstract topological index. It is in fact the dimensionless Hall conductivity $\sigma_{xy}$ itself, provided the Fermi level lies in a gap between magnetic subbands. This was first discovered by Thouless, Kohmoto, Nightingale, and den Nijs, in a seminal paper known by its authors' initials, TKNN[39]. In fact, we've developed the theory here in reverse chronological order. First came TKNN, who found that the contribution $\sigma_{xy}^{(n)}$ to the total Hall conductivity from a band lying entirely below the Fermi level is given by $\sigma_{xy}^{(n)} = \frac{e^2}{h}\, C_n$, where

$$C_n = \frac{i}{2\pi} \int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 \left( \left\langle \frac{\partial \varphi_n}{\partial \theta_1} \middle| \frac{\partial \varphi_n}{\partial \theta_2} \right\rangle - \left\langle \frac{\partial \varphi_n}{\partial \theta_2} \middle| \frac{\partial \varphi_n}{\partial \theta_1} \right\rangle \right) \tag{1.210}$$

is an integral over the Brillouin zone. They proved that this expression is an integer, because

[39] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, *Phys. Rev. Lett.* **49**, 405 (1982).

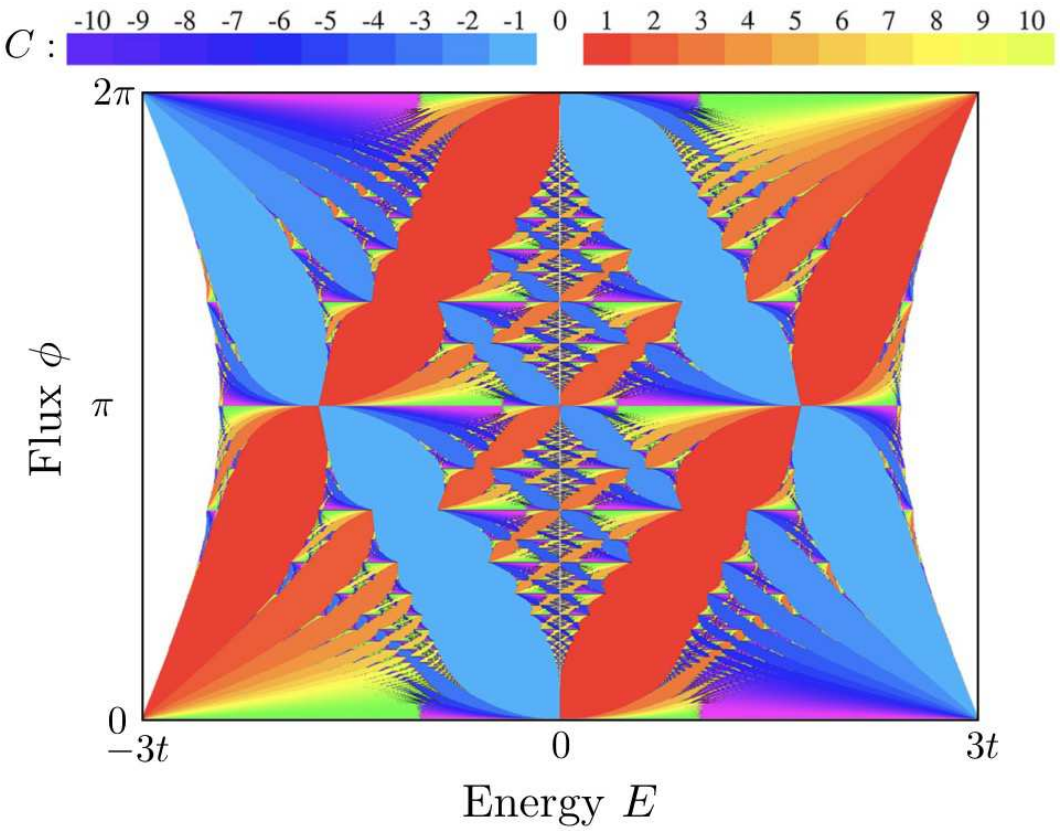


Figure 1.23: Colored Hofstadter butterfly for the honeycomb lattice system, from A. Agazzi, J.-P. Eckman, and G. M. Graf, *J. Stat. Phys.* **156**, 417 (2014).

invoking Stokes' theorem,

$$C_n = \frac{i}{2\pi}\int\limits_0^{2\pi} d\theta_2 \left\langle\, \varphi_n \,\middle|\, \frac{\partial\varphi_n}{\partial\theta_2} \,\right\rangle\Big|_{\theta_1=0}^{\theta_1=2\pi} - \frac{i}{2\pi}\int\limits_0^{2\pi} d\theta_1 \left\langle\, \varphi_n \,\middle|\, \frac{\partial\varphi_n}{\partial\theta_1} \,\right\rangle\Big|_{\theta_2=0}^{\theta_2=2\pi} \quad . \tag{1.211}$$

But since $H(\theta_1,\theta_2)$ is doubly periodic with period $2\pi$ in each of its arguments, we must have

$$\begin{aligned} |\,\varphi_n(\theta_1,2\pi)\,\rangle &= e^{if_n(\theta_1)}\,|\,\varphi_n(\theta_1,0)\,\rangle \\ |\,\varphi_n(2\pi,\theta_2)\,\rangle &= e^{ig_n(\theta_2)}\,|\,\varphi_n(0,\theta_2)\,\rangle \quad . \end{aligned} \tag{1.212}$$

Thus, one finds

$$C_n = \frac{1}{2\pi}\Big(f_n(2\pi) - f_n(0) + g_n(0) - g_n(2\pi)\Big) \quad . \tag{1.213}$$

But we also have

$$\begin{aligned} |\,\varphi_n(0,0)\,\rangle &= e^{-if_n(0)}\,|\,\varphi_n(0,2\pi)\,\rangle = e^{-if_n(0)}\,e^{-ig_n(2\pi)}\,|\,\varphi_n(2\pi,2\pi)\,\rangle \\ &= e^{-if_n(0)}\,e^{-ig_n(2\pi)}\,e^{if_n(2\pi)}\,|\,\varphi_n(2\pi,0)\,\rangle = e^{-if_n(0)}\,e^{-ig_n(2\pi)}\,e^{if_n(2\pi)}\,e^{ig_n(0)}\,|\,\varphi_n(0,0)\,\rangle \quad , \end{aligned} \tag{1.214}$$

and therefore $\exp(2\pi i\, C_n) = 1$ and $C_n \in \mathbb{Z}$. But just as Berry didn't know he had found a holonomy, TKNN didn't know they had found a Chern number. That mathematical feature was first elucidated by Avron, Seiler, and Simon[40], in a paper which is widely appreciated but which, understandably, is not known by its authors' initials.

To see why Hall conductivity is related to Berry curvature, consider an electric field $\boldsymbol{E} = E_y\,\hat{\boldsymbol{y}}$, and the single electron Hamiltonian $H(E_y) = H(0) - eE_y y$, where $H(0) = \frac{\boldsymbol{\pi}^2}{2m} + V(\boldsymbol{r})$ has eigenstates $|\,\alpha\,\rangle$ and eigenvalues $\varepsilon_\alpha$. First order perturbation theory in the electric field term says

$$|\,\alpha'\,\rangle = |\,\alpha\,\rangle - eE_y \sum_\beta{}' \frac{|\,\beta\,\rangle\langle\,\beta\,|\,y\,|\,\alpha\,\rangle}{\varepsilon_\alpha - \varepsilon_\beta} \quad , \tag{1.215}$$

where the prime on the sum means the term with $\beta = \alpha$ is excluded. Let's now compute the expectation of the velocity operator $v_x$ in the perturbed state $|\,\alpha'\,\rangle$. We have, to lowest order,

$$\langle\,\alpha'\,|\,v_x\,|\,\alpha'\,\rangle = -eE_y \sum_\beta{}' \frac{\langle\,\alpha\,|\,v_x\,|\,\beta\,\rangle\langle\,\beta\,|\,y\,|\,\alpha\,\rangle + \langle\,\alpha\,|\,y\,|\,\beta\,\rangle\langle\,\beta\,|\,v_x\,|\,\alpha\,\rangle}{\varepsilon_\alpha - \varepsilon_\beta} \tag{1.216}$$

We now invoke the Feynman-Hellman theorem, which says

$$\langle\,\alpha\,|\,y\,|\,\beta\,\rangle = \frac{\hbar}{i}\,\frac{\langle\,\alpha\,|\,v_y\,|\,\beta\,\rangle}{\varepsilon_\alpha - \varepsilon_\beta} \quad , \tag{1.217}$$

multiply by the electron charge $-e$, divide by the area of the system $\Omega$, and sum using the Fermi distribution over the levels $|\,\alpha\,\rangle$, to obtain the current density $j_x$ :

$$j_x = E_y \cdot \frac{e^2}{h} \cdot \frac{2\pi i\hbar^2}{\Omega} \sum_\alpha \sum_\beta f_\alpha\,(1 - f_\beta)\,\epsilon_{ij}\,\frac{\langle\,\alpha\,|\,v_i\,|\,\beta\,\rangle\langle\,\beta\,|\,v_j\,|\,\alpha\,\rangle}{(\varepsilon_\alpha - \varepsilon_\beta)^2} \quad , \tag{1.218}$$

where $f_\alpha$ is the Fermi function at temperature $T$, chemical potential $\mu$, and energy $\varepsilon_\alpha$. The above expression for $\sigma_{xy} = j_x/E_y$ is known as the *Kubo formula* for the Hall conductivity. At $T = 0$, the Fermi distribution becomes the step function $f_\alpha = \Theta(E_{\mathsf{F}} - E_\alpha)$.

Suppose our system lies on a torus defined by the spatial periods $\boldsymbol{L}_1$ and $\boldsymbol{L}_2$. Define the gauge transformed Hamiltonian

$$\widetilde{H}(\boldsymbol{\theta}) \equiv e^{-i\boldsymbol{q}\cdot\boldsymbol{r}}\,H\,e^{+i\boldsymbol{q}\cdot\boldsymbol{r}} \quad , \tag{1.219}$$

where

$$\boldsymbol{q} = \theta_2\,\frac{\hat{\boldsymbol{z}}\times\boldsymbol{L}_1}{\Omega} - \theta_1\,\frac{\hat{\boldsymbol{z}}\times\boldsymbol{L}_2}{\Omega} \quad , \tag{1.220}$$

with $\Omega = \hat{\boldsymbol{z}}\cdot\boldsymbol{L}_1\times\boldsymbol{L}_2 = 2\pi\ell^2 p$ with $p \in \mathbb{Z}$, *i.e.* the total magnetic flux through the system is an integer multiple of the Dirac quantum. Then

$$\frac{\partial\widetilde{H}}{\partial\theta_i} = \frac{\partial\boldsymbol{q}}{\partial\theta_i}\cdot e^{-i\boldsymbol{q}\cdot\boldsymbol{r}}\,\hbar\boldsymbol{v}\,e^{i\boldsymbol{q}\cdot\boldsymbol{r}} \equiv \frac{\partial\boldsymbol{q}}{\partial\theta_i}\cdot\hbar\widetilde{\boldsymbol{v}} \quad , \tag{1.221}$$

[40] J. E. Avron, R. Seiler, and B. Simon, *Phys. Rev. Lett.* **51**, 51 (1983).

because $[H, \boldsymbol{r}] = (\hbar/i)\,\boldsymbol{v}$. Thus, defining $|\,\widetilde{\alpha}\,\rangle \equiv \exp(-i\boldsymbol{q}\cdot\boldsymbol{r})\,|\,\alpha\,\rangle$, and recalling the definition of the wavevector $\boldsymbol{q} = \epsilon_{ab}\,\theta_a\,\boldsymbol{L}_b \times \hat{\boldsymbol{z}}/\Omega$, we find

$$\frac{\partial\widetilde{H}}{\partial\theta_a} = \hbar\,\epsilon_{ab}\,\epsilon_{ij}\,\frac{\widetilde{v}_i\,L_b^j}{\Omega} \quad . \tag{1.222}$$

We then find

$$\sigma_{xy} = \frac{j_x}{E_y} = \sum_{\alpha\ \mathrm{occ}} \sigma_{xy}^{(\alpha)} \quad , \tag{1.223}$$

where the sum is over occupied states below the Fermi level, and where

$$\sigma_{xy}^{(\alpha)} = \frac{e^2}{h}\cdot 2\pi i \sum_\beta{}' \epsilon_{ij}\,\frac{\langle\,\widetilde{\alpha}\,|\,\frac{\partial\widetilde{H}}{\partial\theta_i}\,|\,\widetilde{\beta}\,\rangle\langle\,\widetilde{\beta}\,|\,\frac{\partial\widetilde{H}}{\partial\theta_j}\,|\,\widetilde{\alpha}\,\rangle}{(\varepsilon_\alpha - \varepsilon_\beta)^2} \quad , \tag{1.224}$$

which is precisely of the form of Eqn. 1.183. Thus, if we now *uniformly average over the boundary phases* $\theta_1$ *and* $\theta_2$, we obtain

$$\big\langle \sigma_{xy}^{(\alpha)} \big\rangle = \frac{e^2}{h}\cdot\frac{i}{2\pi}\int_0^{2\pi} d\theta_1 \int_0^{2\pi} d\theta_2 \sum_\beta{}' \epsilon_{ij}\,\frac{\langle\,\widetilde{\alpha}\,|\,\frac{\partial\widetilde{H}}{\partial\theta_i}\,|\,\widetilde{\beta}\,\rangle\langle\,\widetilde{\beta}\,|\,\frac{\partial\widetilde{H}}{\partial\theta_j}\,|\,\widetilde{\alpha}\,\rangle}{(\varepsilon_\alpha - \varepsilon_\beta)^2} = \frac{e^2}{h}\,C_\alpha \quad , \tag{1.225}$$

*i.e.* each filled band $\alpha$ contributes $\frac{e^2}{h}\,C_\alpha$ to the total Hall conductivity whenever the Fermi level at $T = 0$ lies in a gap between energy bands. For a crystalline (periodic) system, averaging over $\theta_{1,2}$ is tantamount to integrating over the Brillouin zone.

## 1.8 Appendix I : Basis Wavefunctions on a Torus

Periodic boundary conditions on the torus requires $t(\boldsymbol{L}_a)\,|\,\psi\,\rangle = e^{i\theta_a}\,|\,\psi\,\rangle$ for all states $|\,\psi\,\rangle$. Let's examine what this requires for the analytic part $f(z)$. We have

$$\begin{aligned} t(\boldsymbol{L}) &= e^{i\boldsymbol{\kappa}\cdot\boldsymbol{L}/\hbar} = e^{i(\kappa\bar{L}+\kappa^\dagger L)/2\hbar} \\ &= e^{-L\bar{L}/4\ell^2}\,e^{i\kappa\bar{L}/2\hbar}\,e^{i\kappa^\dagger L/2\hbar} \\ &= e^{-L\bar{L}/4\ell^2}\,e^{i\chi}\,e^{z\bar{z}/4\ell^2}\,e^{\bar{L}\bar{\partial}}\,e^{-z\bar{z}/2\ell^2}\,e^{L\partial}\,e^{z\bar{z}/4\ell^2}\,e^{-i\chi} \quad . \end{aligned} \tag{1.226}$$

Thus, with $\psi(\boldsymbol{r}) = e^{i\chi}\,e^{-z\bar{z}/4\ell^2}\,f(z)$, we have

$$\begin{aligned} t(\boldsymbol{L})\,\psi(\boldsymbol{r}) &= e^{-L\bar{L}/4\ell^2}\,e^{i\chi}\,e^{z\bar{z}/4\ell^2}\,e^{\bar{L}\bar{\partial}}\,e^{-z\bar{z}/2\ell^2}\,e^{L\partial}\,f(z) \\ &= e^{i\chi}\,e^{-L\bar{L}/4\ell^2}\,e^{-z\bar{z}/4\ell^2}\,e^{-z\bar{L}/2\ell^2}\,f(z+L) \quad . \end{aligned} \tag{1.227}$$

Thus, we must have

$$f(z+L_a) = e^{i\theta_a}\,e^{L_a\bar{L}_a/4\ell^2}\,e^{z\bar{L}_a/2\ell^2}\,f(z) \quad , \tag{1.228}$$

valid for $a = 1, 2$. Note that integrating the logarithmic derivative of $f(z)$ around the torus yields

$$\begin{aligned}\oint\limits_{\Omega} \frac{dz}{2\pi i}\ \frac{f'(z)}{f(z)} &= \int\limits_0^{L_1} \frac{dz}{2\pi i}\ \frac{d}{dz}\ \ln\biggl(\frac{f(z)}{f(z+L_2)}\biggr) + \int\limits_0^{L_2} \frac{dz}{2\pi i}\ \frac{d}{dz}\ \ln\biggl(\frac{f(z+L_1)}{f(z)}\biggr) \\ &= -\int\limits_0^{L_1} \frac{dz}{2\pi i}\ \frac{\bar{L}_2}{2\ell^2} + \int\limits_0^{L_2} \frac{dz}{2\pi i}\ \frac{\bar{L}_1}{2\ell^2} = \frac{\bar{L}_1 L_2 - L_1 \bar{L}_2}{2\pi i} \cdot \frac{1}{2\ell^2} = \frac{\Omega}{2\pi\ell^2} = N_\phi \quad ,\end{aligned} \tag{1.229}$$

which establishes that $f(z)$ has precisely $N_\phi$ zeros on the torus.

Now consider the Jacobi theta function,

$$\vartheta_1(w\,|\,\sigma) = -i \sum_{n=-\infty}^{\infty} (-1)^n\, e^{i\pi\sigma(n+\frac{1}{2})^2}\, e^{(2n+1)iw} \quad , \tag{1.230}$$

where $\mathsf{Im}\,(\sigma) > 0$. This function is quasiperiodic over the fundamental domain for $w$, which is a parallelogram with sides 1 and $\sigma$, satisfying

$$\begin{aligned}\vartheta_1(w+\pi\,|\,\sigma) &= -\vartheta_1(w\,|\,\sigma) \\ \vartheta_1(w+\pi\sigma\,|\,\sigma) &= -e^{-2iw}\, e^{-i\pi\sigma}\, \vartheta_1(w\,|\,\sigma) \quad .\end{aligned} \tag{1.231}$$

From the above relations, integrating the logarithmic derivative of $\vartheta_1(w\,|\,\sigma)$ establishes that the function has one zero in the fundamental domain, located at $w = 0$. Iterating the second of the above relations, one has

$$\vartheta_1(w + j\pi\sigma\,|\,\sigma) = (-1)^j\, e^{-2ijw}\, e^{-ij^2\pi\sigma}\, \vartheta_1(w\,|\,\sigma) \quad , \tag{1.232}$$

for any integer $j$. In addition, using the Poisson summation formula,

$$\sum_{n=-\infty}^{\infty} \delta(x-n) = \sum_{m=-\infty}^{\infty} e^{2\pi i m x} \quad , \tag{1.233}$$

one derives the identity

$$\vartheta_1(w\,|\,\sigma) = \sqrt{\tfrac{i}{\sigma}}\, e^{-i\pi(1+\sigma)/4}\, e^{-iw^2/\pi\sigma}\, \vartheta\bigl(w\,\big|\,-\tfrac{1}{\sigma}\bigr) \quad . \tag{1.234}$$

Now consider the function

$$f(z) = e^{\bar{L}_1 z^2/4\ell^2 L_1}\, e^{i\lambda z/L_1}\, \vartheta_1\biggl(\frac{\pi z}{L_1} - \pi\zeta\,\bigg|\,\frac{L_2}{N_\phi\, L_1}\biggr) \quad . \tag{1.235}$$

Then

$$\begin{aligned} f(z+L_1) &= -e^{L_1\bar{L}_1/4\ell^2}\,e^{\bar{L}_1 z/2\ell^2}\,e^{i\lambda}\,f(z) \\ f(z+L_2) &= (-1)^{N_\phi}\,e^{L_2\bar{L}_2/4\ell^2}\,e^{\bar{L}_2 z/2\ell^2}\,e^{i\lambda\tau}\,e^{2\pi i N_\phi\zeta}\,f(z) \quad , \end{aligned} \tag{1.236}$$

where $\tau \equiv L_2/L_1$. Invoking the periodicity requirements, we obtain

$$\begin{aligned} e^{i\lambda} &= -e^{i\theta_1} \\ e^{i\lambda\tau}\,e^{2\pi i N_\phi\zeta} &= (-1)^{N_\phi}\,e^{i\theta_2} \quad . \end{aligned} \tag{1.237}$$

Thus, we have

$$\lambda = \theta_1 + (2k_1+1)\pi \tag{1.238}$$

and

$$\zeta = \frac{\theta_2 + \pi N_\phi + 2\pi k_2}{2\pi N_\phi} - \frac{\theta_1 + (2k_1+1)\pi}{2\pi N_\phi}\cdot\tau \quad , \tag{1.239}$$

where $k_1$ and $k_2$ are integers. Since $k_1 \to k_1 + 1$ increases the argument of the $\vartheta$-function by a multiple of $\sigma \equiv \tau/N_\phi$, one can invoke the quasiperiodicity relation, whence one finds that this results in a multiplication of $f(z)$ by a constant. Therefore, we are free to select a fixed value for $k_1$. We choose $k_1 \equiv -1$. Then

$$\begin{aligned} \lambda &= \theta_1 - \pi \\ \zeta &= \frac{\theta_2 + \pi N_\phi + 2\pi k_2}{2\pi N_\phi} + \frac{(\pi-\theta_1)\,\tau}{2\pi N_\phi} \quad . \end{aligned} \tag{1.240}$$

So our basis functions are

$$\psi_k(\boldsymbol{r}) = C\,e^{i\chi}\,e^{i\pi k/N_\phi}\,e^{-z\bar{z}/4\ell^2}\,e^{\bar{L}_1 z^2/4\ell^2 L_1}\,e^{i(\theta_1-\pi)z/L_1}\,\vartheta_1\!\left(\frac{\pi z}{L_1} - \pi\zeta_k \,\middle|\, \frac{L_2}{N_\phi\,L_1}\right) \quad , \tag{1.241}$$

where

$$\zeta_k = \frac{\theta_2 + \pi N_\phi + 2\pi k}{2\pi N_\phi} + \frac{(\pi-\theta_1)\,\tau}{2\pi N_\phi} \tag{1.242}$$

and $C$ is a constant independent of $k$. Then after some work one can show that indeed

$$\begin{aligned} t_1\,\psi_k(\boldsymbol{r}) &= e^{i\theta_1/N_\phi}\,\psi_{k-1}(\boldsymbol{r}) \\ t_2\,\psi_k(\boldsymbol{r}) &= e^{i\theta_2/N_\phi}\,e^{2\pi i k/N_\phi}\,\psi_k(\boldsymbol{r}) \quad . \end{aligned} \tag{1.243}$$

Finally, define

$$w \equiv \frac{z}{L_1} \equiv u + \tau v \quad , \tag{1.244}$$

where $(u,v) \in [0,1] \times [0,1]$. Then with $\phi_x(\boldsymbol{w}) \equiv \psi_k(z = wL_1, \bar{z} = \bar{w}\bar{L}_1)$,

$$\begin{aligned}\phi_k(\boldsymbol{w}) &= C\, e^{i\pi k/N_\phi}\, e^{\pi N_\phi (w-\bar{w})/2\tau_2}\, e^{i(\theta_1-\pi)w}\, \vartheta_1\Big(\pi w - \pi\zeta_k \,\Big|\, \frac{\tau}{N_\phi}\Big) \\ &= C\, e^{i\pi k/N_\phi}\, e^{i\pi N_\phi(uv+\tau v^2)}\, e^{i(\theta_1-\pi)(u+\tau v)}\, \vartheta_1\Big(\pi u - \frac{k\pi}{N_\phi} - \frac{\theta_2+\pi N_\phi}{2N_\phi} + \pi v\tau + \frac{(\theta_1-\pi)\,\tau}{2N_\phi} \,\Big|\, \frac{\tau}{N_\phi}\Big) \quad ,\end{aligned} \tag{1.245}$$

which is holomorphic in $\tau$. The normalization condition is

$$1 = \Omega \int_0^1 du \int_0^1 dv\, \big|\phi_k(u,v)\big|^2 = |C|^2 \left(\frac{2\pi\ell^2 N_\phi^3}{\tau_2}\right)^{1/2} \exp\left(\frac{(\theta_1-\pi)^2\,\tau_2}{2\pi N_\phi}\right) \quad . \tag{1.246}$$

## 1.9 Appendix II : Coherent States and their Path Integral

### 1.9.1 Feynman path integral

The path integral formulation of quantum mechanics is both beautiful and powerful. It is useful in elucidating the quantum-classical correspondence and the semiclassical approximation, in accounting for interference effects, in treatments of tunneling problems via the method of instantons, *etc.* Our goal is to derive and to apply a path integral method for quantum spin. We begin by briefly reviewing the derivation of the usual Feynman path integral.

Consider the propagator $K(x_{\rm i}, x_{\rm f}, T)$, which is the probability amplitude that a particle located at $x = x_{\rm i}$ at time $t = 0$ will be located at $x = x_{\rm f}$ at time $t = T$. We may write

$$\begin{aligned}K(x_{\rm i}, x_{\rm f}, T) &= \langle\, x_{\rm f}\,|\, e^{-iHT/\hbar}\,|\, x_{\rm i}\,\rangle \\ &= \langle\, x_N\,|\, e^{-i\epsilon H/\hbar}\, \mathbf{1}\, e^{-i\epsilon H/\hbar}\, \mathbf{1} \cdots \mathbf{1}\, e^{-i\epsilon H/\hbar}\,|\, x_0\,\rangle\end{aligned} \tag{1.247}$$

where $\epsilon = T/N$, and where we have defined $x_0 \equiv x_{\rm i}$ and $x_N \equiv x_{\rm f}$. We are interested in the limit $N \to \infty$. Inserting $(N-1)$ resolutions of the identity of the form

$$\mathbf{1} = \int_{-\infty}^{\infty} dx_j\,|\, x_j\,\rangle\langle\, x_j\,| \quad , \tag{1.248}$$

we find that we must evaluate matrix elements of the form

$$\begin{aligned}\langle\, x_{j+1}\,|\, e^{-iH\epsilon/\hbar}\,|\, x_j\,\rangle &\approx \int_{-\infty}^{\infty} dp_j\, \langle\, x_{j+1}\,|\, p_j\,\rangle\, \langle\, p_j\,|\, e^{-iT\epsilon/\hbar}\, e^{-iV\epsilon/\hbar}\,|\, x_j\,\rangle \\ &= \int_{-\infty}^{\infty} dp_j\, e^{ip_j(x_{j+1}-x_j)}\, e^{-i\epsilon p_j^2/2m\hbar}\, e^{-i\epsilon V(x_j)/\hbar} \quad .\end{aligned} \tag{1.249}$$

The propagator may now be written as

$$\begin{aligned}\langle\, x_N \,|\, e^{-iHT/\hbar} \,|\, x_0 \,\rangle &\approx \int\limits_{-\infty}^{\infty} \prod_{j=1}^{N-1} dx_j \int\limits_{-\infty}^{\infty} \prod_{k=0}^{N-1} dp_k \, \exp\left\{ i \sum_{k=0}^{N-1} \left[ p_k(x_{k+1} - x_k) - \frac{\epsilon\, p_k^2}{2m\hbar} - \frac{\epsilon}{\hbar}\, V(x_k) \right] \right\} \\ &= \left( \frac{2\pi\hbar m}{i\epsilon} \right)^N \int\limits_{-\infty}^{\infty} \prod_{j=1}^{N-1} dx_j \, \exp\left\{ \frac{i\epsilon}{\hbar} \sum_{k=1}^{N-1} \left[ \tfrac{1}{2} m \Big( \frac{x_{j+1} - x_j}{\epsilon} \Big)^2 - V(x_j) \right] \right\} \\ &\equiv \int\limits_{\substack{x(0)=x_{\rm i} \\ x(T)=x_{\rm f}}} \mathcal{D}x(t) \, \exp\left\{ \frac{i}{\hbar} \int_0^T dt \left[ \tfrac{1}{2} m\dot{x}^2 - V(x) \right] \right\} \quad , \end{aligned} \tag{1.250}$$

where we absorb the prefactor into the measure $\mathcal{D}x(t)$. Note the boundary conditions on the path integral at $t = 0$ and $t = T$. In the semiclassical approximation, we assume that the path integral is dominated by trajectories $x(t)$ which extremize the argument of the exponential in the last term above. This quantity is (somewhat incorrectly) identified as the classical action, $\mathcal{S}$, and the action-extremizing equations are of course the Euler-Lagrange equations. Setting $\delta\mathcal{S} = 0$ yields Newton's second law, $m\ddot{x} = -\partial V/\partial x$, which is to be solved subject to the two boundary conditions.

The imaginary time version, which yields the thermal propagator, is obtained by writing $T = -i\hbar\beta$ and $t = -i\tau$, in which case

$$\langle\, x_{\rm f} \,|\, e^{-\beta H} \,|\, x_{\rm i} \,\rangle = \int\limits_{\substack{x(0)=x_{\rm i} \\ x(\hbar\beta)=x_{\rm f}}} \mathcal{D}x(\tau) \, \exp\left\{ -\frac{1}{\hbar} \overbrace{\int_0^{\hbar\beta} d\tau \left[ \tfrac{1}{2} m\dot{x}^2 + V(x) \right]}^{\text{Euclidean action } \mathcal{S}_{\rm E}} \right\} \quad . \tag{1.251}$$

The partition function is the trace of the thermal propagator, *viz.*

$$Z = \mathsf{Tr}\; e^{-\beta H} = \int\limits_{-\infty}^{\infty} dx \, \langle\, x \,|\, e^{-\beta H} \,|\, x \,\rangle = \int\limits_{x(0)=x(\hbar\beta)} \mathcal{D}x(\tau) \, \exp\big( -\mathcal{S}_{\rm E}[x(\tau)]/\hbar \big) \tag{1.252}$$

The equations of motion derived from $\mathcal{S}_{\rm E}$ are $m\ddot{x} = +\partial V/\partial x$, corresponding to motion in the 'inverted potential'.

### 1.9.2 Primer on coherent states

We now turn to the method of coherent state path integration. In order to discuss this, we must first introduce the notion of coherent states. This is most simply done by appealing to the

one-dimensional simple harmonic oscillator,

$$H = \frac{p^2}{2m} + \tfrac{1}{2}m\omega_0^2 x^2 = \hbar\omega_0\,(a^\dagger a + \tfrac{1}{2}) \tag{1.253}$$

where $a$ and $a^\dagger$ are ladder operators,

$$a = \ell\,\partial_x + \frac{x}{2\ell} \qquad , \qquad a^\dagger = -\ell\,\partial_x + \frac{x}{2\ell} \tag{1.254}$$

with $\ell \equiv \sqrt{\hbar/2m\omega_0}$. *Exercise: Check that* $[a, a^\dagger] = 1$.

The ground state satisfies $a\,\psi_0(x) = 0$, which yields

$$\psi_0(x) = (2\pi\ell^2)^{-1/4}\,\exp(-x^2/4\ell^2) \quad . \tag{1.255}$$

The normalized coherent state $|\,z\,\rangle$ is defined as

$$|\,z\,\rangle = e^{-\frac{1}{2}|z|^2}\,e^{za^\dagger}|\,0\,\rangle = e^{-\frac{1}{2}|z|^2}\sum_{n=0}^{\infty}\frac{z^n}{\sqrt{n!}}\,|\,n\,\rangle \quad . \tag{1.256}$$

The coherent state is an eigenstate of the annihilation operator $a$:

$$a\,|\,z\,\rangle = z\,|\,z\,\rangle \qquad \Longleftrightarrow \qquad \langle\,z\,|\,a^\dagger = \langle\,z\,|\,\bar{z} \quad . \tag{1.257}$$

The overlap of coherent states is given by

$$\langle\,z_1\,|\,z_2\,\rangle = e^{-\frac{1}{2}|z_1|^2}\,e^{-\frac{1}{2}|z_2|^2}\,e^{\bar{z}_1 z_2} \quad , \tag{1.258}$$

hence different coherent states are not orthogonal. Despite this nonorthogonality, the coherent states allow a simple resolution of the identity,

$$\mathbf{1} = \int\frac{d^2z}{2\pi i}\,|\,z\,\rangle\langle\,z\,| \qquad , \qquad \frac{d^2z}{2\pi i} \equiv \frac{d\,\mathrm{Re}\,z\;d\,\mathrm{Im}\,z}{\pi} \tag{1.259}$$

which is straightforward to establish.

To gain some physical intuition about the coherent states, define

$$z \equiv \frac{Q}{2\ell} + \frac{i\ell P}{\hbar} \quad . \tag{1.260}$$

One finds (*exercise!*)

$$\psi_{P,Q}(x) = \langle\,x\,|\,z\,\rangle = (2\pi\ell^2)^{-1/4}\,e^{-iPQ/2\hbar}\,e^{iPx/\hbar}\,e^{-(x-Q)^2/4\ell^2} \quad , \tag{1.261}$$

hence the coherent state $\psi_{P,Q}(x)$ is a wavepacket Gaussianly localized about $x = Q$, but oscillating with momentum $P$.

For example, we can compute

$$\begin{aligned}\langle\, Q,P\,|\,q\,|\,Q,P\,\rangle &= \langle\, z\,|\,\ell\,(a+a^\dagger)\,|\,z\,\rangle = 2\ell\ \mathsf{Re}\ z = Q \\ \langle\, Q,P\,|\,p\,|\,Q,P\,\rangle &= \langle\, z\,|\,\frac{\hbar}{2i\ell}\,(a-a^\dagger)\,|\,z\,\rangle = \frac{\hbar}{\ell}\,\mathsf{Im}\ z = P\end{aligned} \tag{1.262}$$

as well as

$$\begin{aligned}\langle\, Q,P\,|\,q^2\,|\,Q,P\,\rangle &= \langle\, z\,|\,\ell^2\,(a+a^\dagger)^2\,|\,z\,\rangle = Q^2+\ell^2 \\ \langle\, Q,P\,|\,p^2\,|\,Q,P\,\rangle &= -\langle\, z\,|\,\frac{\hbar^2}{4\ell^2}(a-a^\dagger)^2\,|\,z\,\rangle = P^2+\frac{\hbar^2}{4\ell^2}\quad .\end{aligned} \tag{1.263}$$

Thus, the root mean square fluctuations in the coherent state $|\,Q,P\,\rangle$ are

$$\Delta q = \ell = \sqrt{\frac{\hbar}{2m\omega_0}}\qquad,\qquad \Delta p = \frac{\hbar}{2\ell} = \sqrt{\frac{m\hbar\omega_0}{2}}\quad, \tag{1.264}$$

and $\Delta q\cdot\Delta p = \frac{1}{2}\,\hbar$. Thus we learn that the coherent state $\psi_{Q,P}(q)$ is localized in phase space, *i.e.* in both position and momentum. If we have a general operator $\hat{A}(q,p)$, we can then write

$$\langle\, Q,P\,|\,\hat{A}(q,p)\,|\,Q,P\,\rangle = A(Q,P) + \mathcal{O}(\hbar)\quad, \tag{1.265}$$

where $A(Q,P)$ is formed from $\hat{A}(q,p)$ by replacing $q\to Q$ and $p\to P$. Since

$$\frac{d^2z}{2\pi i} \equiv \frac{d\,\mathsf{Re}\,z\,d\,\mathsf{Im}\,z}{\pi} = \frac{dQ\,dP}{2\pi\hbar}\quad, \tag{1.266}$$

we can write the trace using coherent states as

$$\mathsf{Tr}\,\hat{A} = \frac{1}{2\pi\hbar}\int\limits_{-\infty}^{\infty}\!dQ\int\limits_{-\infty}^{\infty}\!dP\,\langle\, Q,P\,|\,\hat{A}\,|\,Q,P\,\rangle\quad. \tag{1.267}$$

We now can understand the origin of the factor $2\pi\hbar$ in the denominator of each $(q_i,p_i)$ integral over classical phase space in the metric $d\mu = \prod_i \frac{dq_i dp_i}{2\pi\hbar}$.

Note that $\omega_0$ is arbitrary in our discussion. By increasing $\omega_0$ , the states become more localized in $q$ and more plane wave like in $p$. However, so long as $\omega_0$ is finite, the width of the coherent state in each direction is proportional to $\hbar^{1/2}$, and thus vanishes in the classical limit.

The resolution of the identity in Eqn. 1.259 prompts the following question. Suppose we consider an infinite discrete lattice $\{|\,z_{m,n}\,\rangle\}$ of coherent states, with $z_{m,n} = (m+in)\sqrt{\pi}$, called a *von Neumann lattice* of coherent states. The dimensionless phase space area $A$ per unit cell of this lattice is $\pi$, which is the denominator in the integral resolution of the identity in Eqn. 1.259. One might expect, then, that while this basis is not orthogonal, since

$$\langle\, z_{m,n}\,|\,z_{m',n'}\,\rangle = e^{-\pi(m-m')^2/2}\,e^{-\pi(n-n')^2/2}\,e^{i\pi(mn'-nm')}\quad, \tag{1.268}$$

that the Gaussian overlaps could be undone, *i.e.* the overlap matrix could be inverted, and a complete and orthonormal set of *localized* LLL wavefunctions could be constructed. Alas, this is impossible! As shown by Perelomov[41], there is a single linear dependence relation among the von Neumann lattice of coherent states, when $A = \pi$. For $S < \pi$, the lattice is overcomplete by a finite amount per unit area. Similarly, when $A > \pi$, the lattice is undercomplete by a finite amount per unit area. But when $A = \pi$, it is undercomplete by precisely *one state*. This state is necessary to include in order for the filled Landau level to carry a Chern number $C = 1$. Thus, the von Neumann lattice cannot be used as a basis to describe the quantum Hall effect.

### 1.9.3 Coherent state path integral

Now we derive the imaginary time path integral. We write

$$\langle\, z_{\rm f}\,|\, e^{-\beta H}\,|\, z_{\rm i}\,\rangle = \langle\, z_N\,|\, e^{-\epsilon H/\hbar}\, \mathbf{1}\, e^{-\epsilon H/\hbar} \cdots \mathbf{1}\, e^{-\epsilon H/\hbar}\,|\, z_0\,\rangle \quad , \tag{1.269}$$

inserting resolutions of the identity at $N-1$ points, as before. We next evaluate the matrix element

$$\begin{aligned}\langle\, z_j\,|\, e^{-\epsilon H/\hbar}\,|\, z_{j-1}\,\rangle &= \langle\, z_j\,|\, z_{j-1}\,\rangle \cdot \left\{ 1 - \frac{\epsilon}{\hbar}\,\frac{\langle\, z_j\,|\, H\,|\, z_{j-1}\,\rangle}{\langle\, z_j\,|\, z_{j-1}\,\rangle} + \dots \right\} \\ &\simeq \langle\, z_j\,|\, z_{j-1}\,\rangle\, \exp\left\{ -\frac{\epsilon}{\hbar}\, H(\bar z_j\,|z_{j-1}) \right\}\end{aligned} \tag{1.270}$$

where

$$H(\bar z\,|w) \equiv \frac{\langle\, z\,|\, H\,|\, w\,\rangle}{\langle\, z\,|\, w\,\rangle} = e^{-\bar z w}\, \langle\, 0\,|\, e^{\bar z a}\, H(a^\dagger, a)\, e^{w a^\dagger}\,|\, 0\,\rangle \quad . \tag{1.271}$$

This last equation is extremely handy. It says, upon invoking eqn. 1.257, that if $H(a, a^\dagger)$ is *normal ordered* such that all creation operators $a^\dagger$ appear to the *left* of all destruction operators $a$, then $H(\bar z|w)$ is obtained from $H(a^\dagger, a)$ simply by sending $a^\dagger \to \bar z$ and $a \to w$. This is because $a$ acting to the right on $|\, w\,\rangle$ yields its eigenvalue $w$, while $a^\dagger$ acting to the left on $\langle\, z\,|$ generates $\bar z$. Note that the function $H(\bar z|w)$ is holomorphic in $w$ and in $\bar z$, but is completely independent of their complex conjugates $\bar w$ and $z$.

The overlap between coherent states at consecutive time slices may be written

$$\langle\, z_j\,|\, z_{j-1}\,\rangle = \exp\left\{ -\tfrac{1}{2}\Big[\bar z_j (z_j - z_{j-1}) - z_{j-1}(\bar z_j - \bar z_{j-1})\Big] \right\} \quad , \tag{1.272}$$

[41] A. M. Perelomov, *Theor. Math. Phys.* **6**, 156 (1971).

hence

$$\begin{aligned}\langle\, z_N \,|\, z_{N-1} \,\rangle \cdots \langle\, z_1 \,|\, z_0 \,\rangle = \exp\left\{ \tfrac{1}{2} \sum_{j=1}^{N-1} \Big[ z_j(\bar z_{j+1} - \bar z_j) - \bar z_j(z_j - z_{j-1}) \Big] \right\} \\ \times \exp\left\{ \tfrac{1}{2}\, z_0(\bar z_1 - \bar z_0) - \tfrac{1}{2}\, \bar z_N(z_N - z_{N-1}) \right\}\end{aligned} \tag{1.273}$$

which allows us to write down the path integral expression for the propagator,

$$\begin{aligned}\langle\, z_{\rm f} \,|\, e^{-\beta H} \,|\, z_{\rm i} \,\rangle &= \int \prod_{j=1}^{N-1} \frac{d^2 z_j}{2\pi i}\, \exp\Big( - \mathcal{S}_{\rm E}[\{z_j, \bar z_j\}]/\hbar \Big) \\ \mathcal{S}_{\rm E}[\{z_j, \bar z_j\}]/\hbar &= \sum_{j=1}^{N-1} \Big[ \tfrac{1}{2}\, \bar z_j(z_j - z_{j-1}) - \tfrac{1}{2}\, z_j(\bar z_{j+1} - \bar z_j) \Big] + \frac{\epsilon}{\hbar} \sum_{j=1}^{N} H(\bar z_j \,|z_{j-1}) \\ &\qquad + \tfrac{1}{2}\, \bar z_{\rm f}\big(z_{\rm f} - z_{N-1}\big) - \tfrac{1}{2}\, z_{\rm i}\big(\bar z_1 - \bar z_{\rm i}\big) \quad .\end{aligned} \tag{1.274}$$

In the limit $N \to \infty$, we identify the continuum Euclidean action

$$\begin{aligned}\mathcal{S}_{\rm E}[\{z(\tau), \bar z(\tau)\}]/\hbar = \int_0^{\hbar\beta} d\tau \left\{ \frac{1}{2}\left( \bar z\, \frac{dz}{d\tau} - z\, \frac{d\bar z}{d\tau} \right) + \frac{1}{\hbar}\, H(\bar z \,|z) \right\} \\ + \tfrac{1}{2}\, \bar z_{\rm f}\big[ z_{\rm f} - z(\hbar\beta) \big] - \tfrac{1}{2}\, z_{\rm i}\big[ \bar z(0) - \bar z_{\rm i} \big]\end{aligned} \tag{1.275}$$

and write the continuum expression for the path integral,

$$\langle\, z_{\rm f} \,|\, e^{-\beta H} \,|\, z_{\rm i} \,\rangle = \int\limits_{\substack{z(0)=z_{\rm i} \\ \bar z(\hbar\beta)=\bar z_{\rm f}}} \mathcal{D}[z(\tau), \bar z(\tau)]\, e^{-\mathcal{S}_{\rm E}[\{z(\tau), \bar z(\tau)\}]/\hbar} \quad . \tag{1.276}$$

The corresponding real time expression is given by

$$\langle\, z_{\rm f} \,|\, e^{-iHT/\hbar} \,|\, z_{\rm i} \,\rangle = \int\limits_{\substack{z(0)=z_{\rm i} \\ \bar z(T)=\bar z_{\rm f}}} \mathcal{D}[z(t), \bar z(t)]\, e^{i\mathcal{S}[\{z(t), \bar z(t)\}]/\hbar} \tag{1.277}$$

with

$$\begin{aligned}\mathcal{S}[\{z(t), \bar z(t)\}]/\hbar = \int_0^{T} dt \left\{ \frac{1}{2i}\left( z\, \frac{d\bar z}{dt} - \bar z\, \frac{dz}{dt} \right) - \frac{1}{\hbar}\, H(\bar z \,|z) \right\} \\ + \tfrac{1}{2}\, i\bar z_{\rm f}\big[ z_{\rm f} - z(T) \big] - \tfrac{1}{2}\, i z_{\rm i}\big[ \bar z(0) - \bar z_{\rm i} \big] \quad .\end{aligned} \tag{1.278}$$

The continuum limit is in a sense justified by examining the discrete equations of motion,

$$\begin{aligned}\frac{1}{\hbar}\frac{\partial \mathcal{S}_{\rm E}}{\partial z_k} &= \bar z_k - \bar z_{k+1} + \frac{\epsilon}{\hbar}\,\frac{\partial H(\bar z_{k+1}|z_k)}{\partial z_k} \\ \frac{1}{\hbar}\frac{\partial \mathcal{S}_{\rm E}}{\partial \bar z_k} &= z_k - z_{k-1} + \frac{\epsilon}{\hbar}\,\frac{\partial H(\bar z_k|z_{k-1})}{\partial \bar z_k} \quad ,\end{aligned} \tag{1.279}$$

which have the sensible continuum limit

$$\hbar\,\frac{d\bar z}{d\tau} = \frac{\partial H(\bar z\,|z)}{\partial z} \qquad , \qquad \hbar\,\frac{dz}{d\tau} = -\frac{\partial H(\bar z\,|z)}{\partial \bar z} \tag{1.280}$$

with boundary conditions $\bar z(\hbar\beta) = \bar z_{\rm f}$ and $z(0) = z_{\rm i}$. Note that there are only two boundary conditions – one on $z(0)$, the other on $\bar z(\hbar\beta)$. *The function $z(\tau)$ (or its discrete version $z_j$) is evolved forward from initial data $z_{\rm i}$, while $\bar z(\tau)$ (or $\bar z_j$) is evolved backward from final data $\bar z_{\rm f}$.* This is the proper number of boundary conditions to place on two first order differential (or finite difference) equations. It is noteworthy that the action of eqn. 1.274 or eqn. 1.275 imposes only a *finite* penalty on *discontinuous* paths.[42] Nevertheless, the paths which extremize the action are continuous throughout the interval $\tau \in (0, \hbar\beta)$. As $z(\tau)$ is integrated forward from $z_{\rm i}$, its final value $z(\hbar\beta)$ will in general be different from $z_{\rm f}$. Similarly, $\bar z(\tau)$ integrated backward from $\bar z_{\rm f}$ will in general yield an endpoint value $\bar z(0)$ which differs from $\bar z_{\rm i}$. The differences $z(\hbar\beta) - z_{\rm f}$ and $\bar z(0) - \bar z_{\rm i}$ are often identified as path discontinuities, but the fact is that the equations of motion know nothing about either $z_{\rm f}$ or $\bar z_{\rm i}$. These difference terms do enter in a careful accounting of the action formulae of eqns. 1.274 and 1.275, however.

The importance of the boundary terms is nicely illustrated in a computation of the semiclassical imaginary time propagator for the harmonic oscillator. With $H = \hbar\omega_0\, a^\dagger a$ (dropping the constant term for convenience), we have

$$\begin{aligned}\langle\, z_{\rm f}\,|\,\exp(-\beta\hbar\omega_0\, a^\dagger a)\,|\,z_{\rm i}\,\rangle &= e^{-\frac12|z_{\rm f}|^2-\frac12|z_{\rm i}|^2}\sum_{m,n=0}^{\infty}\frac{\bar z_{\rm f}^{\,m}\, z_{\rm i}^{\,n}}{\sqrt{m!\,n!}}\,\langle\, m\,|\,\exp(-\beta\hbar\omega_0\, a^\dagger a)\,|\,n\,\rangle \\ &= \exp\left\{-\tfrac12|z_{\rm f}|^2 - \tfrac12|z_{\rm i}|^2 + \bar z_{\rm f} z_{\rm i}\, e^{-\beta\hbar\omega_0}\right\}\end{aligned} \tag{1.281}$$

The Euclidean action is $L_{\rm E} = \frac12\hbar(\bar z\dot z - z\dot{\bar z}) + \hbar\omega_0\,\bar z z$, so the equations of motion are

$$\hbar\dot{\bar z} = \frac{\partial H}{\partial z} = \hbar\omega_0\,\bar z \qquad , \qquad \hbar\dot z = -\frac{\partial H}{\partial \bar z} = -\hbar\omega_0\, z \tag{1.282}$$

subject to boundary conditions $z(0) = z_{\rm i}$, $\bar z(\hbar\beta) = \bar z_{\rm f}$. The solution is

$$z(\tau) = z_{\rm i}\, e^{-\omega_0\tau} \qquad , \qquad \bar z(\tau) = \bar z_{\rm f}\, e^{\omega_0(\tau-\hbar\beta)} \quad . \tag{1.283}$$

[42] In the Feynman path integral, discontinuous paths contribute an infinite amount to the action, and are therefore suppressed.

Along the 'classical path' the Euclidean Lagrangian vanishes: $L_{\rm E} = 0$. The entire contribution to the action therefore comes from the boundary terms:

$$\begin{aligned}\mathcal{S}_{\rm E}^{\rm cl}/\hbar &= 0 + \tfrac{1}{2}\bar{z}_{\rm f}(z_{\rm f} - z_{\rm i}\,e^{-\beta\hbar\omega_0}) - \tfrac{1}{2}z_i(\bar{z}_{\rm f}\,e^{-\beta\hbar\omega_0} - \bar{z}_{\rm i}) \\ &= \tfrac{1}{2}|z_{\rm f}|^2 + \tfrac{1}{2}|z_{\rm i}|^2 - \bar{z}_{\rm f}z_{\rm i}\,e^{-\beta\hbar\omega_0} \quad ,\end{aligned} \tag{1.284}$$

What remains is to compute the fluctuation determinant. We write

$$\begin{aligned}z_j &= z_j^{\rm cl} + \eta_j \\ \bar{z}_j &= \bar{z}_j^{\rm cl} + \bar{\eta}_j\end{aligned} \tag{1.285}$$

and expand the action as

$$\begin{aligned}\mathcal{S}_{\rm E}[\{z_j,\bar{z}_j\}] &= \mathcal{S}_{\rm E}[\{z_j^{\rm cl},\bar{z}_j^{\rm cl}\}] + \frac{\partial^2\mathcal{S}_{\rm E}}{\partial\bar{z}_i\partial z_j}\,\bar{\eta}_i\eta_j + \frac{1}{2}\,\frac{\partial^2\mathcal{S}_{\rm E}}{\partial z_i\partial z_j}\,\eta_i\eta_j + \frac{1}{2}\,\frac{\partial^2\mathcal{S}_{\rm E}}{\partial\bar{z}_i\partial\bar{z}_j}\,\bar{\eta}_i\bar{\eta}_j + \ldots \\ &\equiv \mathcal{S}_{\rm E}^{\rm cl} + \frac{\hbar}{2}\begin{pmatrix}\bar{z}_i & z_i\end{pmatrix}\begin{pmatrix}A_{ij} & B_{ij}\\ C_{ij} & A_{ij}^{\rm t}\end{pmatrix}\begin{pmatrix}z_j\\ \bar{z}_j\end{pmatrix} + \ldots\end{aligned} \tag{1.286}$$

For general $H$, we obtain

$$\begin{aligned}A_{ij} &= \delta_{ij} - \delta_{i,j+1} + \frac{\epsilon}{\hbar}\,\frac{\partial^2 H(\bar{z}_i\,|z_j)}{\partial\bar{z}_i\,\partial z_j}\,\delta_{i,j+1} \\ B_{ij} &= \frac{\epsilon}{\hbar}\,\frac{\partial^2 H(\bar{z}_i\,|z_{i-1})}{\partial\bar{z}_i^2}\,\delta_{i,j} \\ C_{ij} &= \frac{\epsilon}{\hbar}\,\frac{\partial^2 H(\bar{z}_{i+1}|z_i)}{\partial z_i^2}\,\delta_{i,j}\end{aligned} \tag{1.287}$$

with $i$ and $j$ running from 1 to $N-1$. The contribution of the fluctuation determinant to the matrix element is then

$$\begin{aligned}\int\prod_{j=1}^{N-1}\frac{d^2\eta_i}{2\pi i}\,\exp\Bigg\{ &-\frac{1}{2}\begin{pmatrix}\mathsf{Re}\;\eta_k & \mathsf{Im}\;\eta_k\end{pmatrix}\begin{pmatrix}1 & 1\\ -i & i\end{pmatrix}\begin{pmatrix}A_{kl} & B_{kl}\\ C_{kl} & A_{lk}\end{pmatrix}\begin{pmatrix}1 & i\\ 1 & -i\end{pmatrix}\begin{pmatrix}\mathsf{Re}\;\eta_l\\ \mathsf{Im}\;\eta_l\end{pmatrix}\Bigg\} \\ &= \mathsf{det}^{-1/2}\begin{pmatrix}A & B\\ C & A^{\rm t}\end{pmatrix}\end{aligned}$$

In the case of the harmonic oscillator discussed above, we have $B_{ij} = C_{ij} = 0$, and since $A_{ij}$ has no elements above its diagonal and $A_{ii} = 1$ for all $i$, we simply have that the determinant contribution is unity.

## 1.10 Appendix III : Gauss-Bonnet and Pontrjagin

### 1.10.1 Gauss-Bonnet theorem

There is a deep result in mathematics, the Gauss-Bonnet theorem, which connects the *local geometry* of a two-dimensional manifold to its *global topology*. The content of the theorem is as follows:

$$\int\limits_{\mathcal{M}} dS\; K = 2\pi\,\chi(\mathcal{M}) = 2\pi \sum_i \mathsf{ind}_{\boldsymbol{x}_i}(\boldsymbol{V})\;, \tag{1.288}$$

where $\mathcal{M}$ is a 2-manifold (a topological space locally homeomorphic to $\mathbb{R}^2$), $K$ is the local *Gaussian curvature* of $\mathcal{M}$, given by $K = (R_1\,R_2)^{-1}$, where $R_{1,2}$ are the principal radii of curvature at a given point, and $dS$ is the differential area element. Here $\boldsymbol{V}(\boldsymbol{x})$ is a vector field on $\mathcal{M}$, and $\mathsf{ind}_{\boldsymbol{x}_i}(\boldsymbol{V})$ refers to the *index* of $\boldsymbol{V}$ at its $i^{\rm th}$ singularity $\boldsymbol{x}_i$. The index is in general defined relative to any closed curve in $\mathcal{M}$, and is given by the winding number of $\boldsymbol{V}(\boldsymbol{x})$ around the curve, *viz.*

$$\mathsf{ind}_{\mathcal{C}}(\boldsymbol{V}) = \oint\limits_{\mathcal{C}} d\boldsymbol{x}\cdot\boldsymbol{\nabla}\tan^{-1}\!\left(\frac{V_2(\boldsymbol{x})}{V_1(\boldsymbol{x})}\right)\quad . \tag{1.289}$$

If $\mathcal{C}$ encloses no singularities, then the index necessarily vanishes, but if $\mathcal{C}$ encloses one or more singularities, the index is an integer, given by the winding number of $\boldsymbol{V}$ around the curve $\mathcal{C}$.

The quantity $\chi(\mathcal{M})$ is called the *Euler characteristic* of $\mathcal{M}$ and is given by $\chi(\mathcal{M}) = 2-2g$, where $g$ is the *genus* of $\mathcal{M}$, which is the number of holes (or handles) of $\mathcal{M}$. Furthermore, $\boldsymbol{V}(\boldsymbol{x})$ can be *any* smooth vector field on $\mathcal{M}$, with $\boldsymbol{x}_i$ the singularity points of that vector field[43].

To apprehend the content of the Gauss-Bonnet theorem, it is helpful to consider an example. Let $\mathcal{M} = \mathbb{S}^2$ be the unit 2-sphere, as depicted in fig. 1.24. At any point on the unit 2-sphere, the radii of curvature are degenerate and both equal to $R = 1$, hence $K = 1$. If we integrate the Gaussian curvature over the sphere, we thus get $4\pi = 2\pi\,\chi\big(\mathbb{S}^2\big)$, which says $\chi(\mathbb{S}^2) = 2-2g = 2$, which agrees with $g = 0$ for the sphere. Furthermore, the Gauss-Bonnet theorem says that *any* smooth vector field on $\mathbb{S}^2$ *must* have a singularity or singularities, with the total index summed over the singularities equal to $+2$. The vector field sketched in the left panel of fig. 1.24 has two index $+1$ singularities, which could be taken at the north and south poles, but which could be anywhere. Another possibility, depicted in the right panel of fig. 1.24, is that there is a one singularity with index $+2$.

In fig. 1.25 we show examples of manifolds with genii $g = 1$ and $g = 2$. The case $g = 1$ is the familiar 2-torus, which is topologically equivalent to a product of circles: $\mathbb{T}^2 \cong \mathbb{S}^1 \times \mathbb{S}^1$, and is thus coordinatized by two angles $\theta_1$ and $\theta_2$. A smooth vector field pointing in the direction of increasing $\theta_1$ never vanishes, and thus has no singularities, consistent with $g = 1$ and $\chi\big(\mathbb{T}^2\big) = 0$.

[43] The singularities $\boldsymbol{x}_i$ are fixed points of the dynamical system $\dot{\boldsymbol{x}} = \boldsymbol{V}(\boldsymbol{x})$.

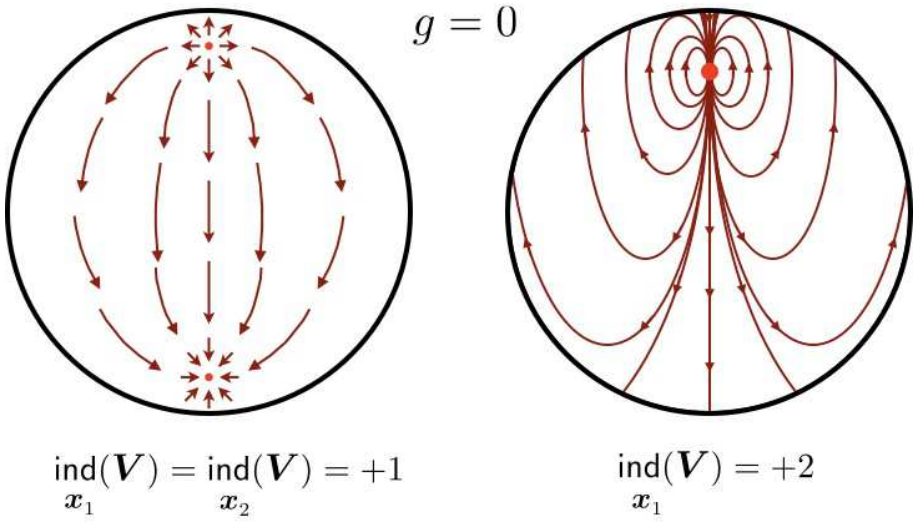


Figure 1.24: Two smooth vector fields on the sphere $\mathbb{S}^2$, which has genus $g = 0$. Left panel: two index $+1$ singularities. Right panel: one index $+2$ singularity.

Topologically, one can define a torus as the quotient space $\mathbb{R}^2/\mathbb{Z}^2$, or as a square with opposite sides identified. This is what mathematicians call a 'flat torus' – one with curvature $K = 0$ everywhere. Of course, such a torus cannot be embedded in three-dimensional Euclidean space; a two-dimensional figure embedded in a three-dimensional Euclidean space inherits a metric due to the embedding, and for a physical torus, like the surface of a bagel, the Gaussian curvature is only zero *on average*.

The $g = 2$ surface $\mathcal{M}$ shown in the right panel of fig. 1.25 has Euler characteristic $\chi(\mathcal{M}) = -2$, which means that any smooth vector field on $\mathcal{M}$ must have singularities with indices totalling $-2$. One possibility, depicted in the figure, is to have two saddle points with index $-1$; one of these singularities is shown in the figure (the other would be on the opposite side).

### 1.10.2 The Pontrjagin index

Consider an $N$-dimensional vector field $\dot{\boldsymbol{x}} = \boldsymbol{V}(\boldsymbol{x})$, and let $\hat{\boldsymbol{n}}(\boldsymbol{x})$ be the unit vector field defined by $\hat{\boldsymbol{n}}(\boldsymbol{x}) = \boldsymbol{V}(\boldsymbol{x})/|\boldsymbol{V}(\boldsymbol{x})|$. Consider now a unit sphere in $\hat{\boldsymbol{n}}$ space, which is of dimension $(N-1)$. If we integrate over this surface, we obtain

$$\Omega_N = \oint d\sigma_a \, n^a = \frac{2\pi^{(N-1)/2}}{\Gamma\left(\frac{N-1}{2}\right)} \,, \tag{1.290}$$

which is the surface area of the unit sphere $\mathbb{S}^{N-1}$. Thus, $\Omega_2 = 2\pi$, $\Omega_3 = 4\pi$, $\Omega_4 = 2\pi^2$, *etc.*

Now consider a change of variables to those over the surface of the sphere, $(\xi_1, \ldots, \xi_{N-1})$. We

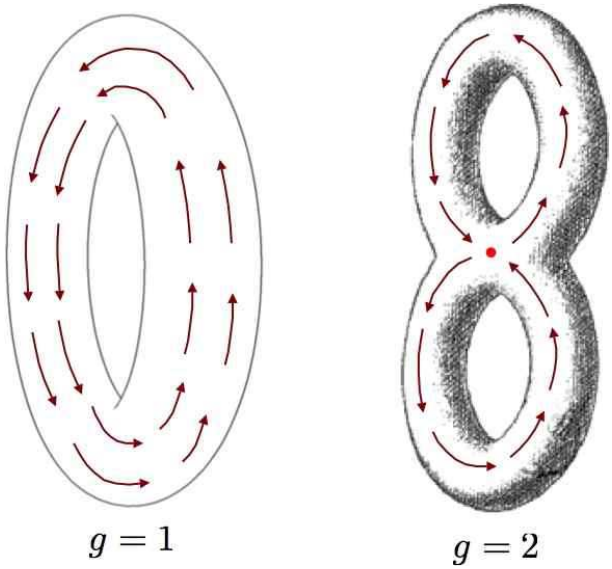


Figure 1.25: Smooth vector fields on the torus $\mathbb{T}^2$, and on a 2-manifold $\mathcal{M}$ of genus $g = 2$

then have

$$\Omega_N = \oint\limits_{\mathbb{S}^{N-1}} d\sigma_a \, n^a = \oint d^{N-1}\xi \, \epsilon_{a_1 \cdots a_N} \, n^{a_1} \, \frac{\partial n^{a_2}}{\partial \xi_1} \cdots \frac{\partial n^{a_N}}{\partial \xi_{N-1}} \tag{1.291}$$

The topological charge is then

$$Q = \frac{1}{\Omega_N} \oint d^{N-1}\xi \, \epsilon_{a_1 \cdots a_N} \, n^{a_1} \, \frac{\partial n^{a_2}}{\partial \xi_1} \cdots \frac{\partial n^{a_N}}{\partial \xi_{N-1}} \tag{1.292}$$

The quantity $Q$ is an *integer topological invariant* which characterizes the map from the surface $(\xi_1, \ldots, \xi_{N-1})$ to the unit sphere $|\hat{\boldsymbol{n}}| = 1$. In mathematical parlance, $Q$ is known as the *Pontrjagin index* of this map.

This analytical development recapitulates some basic topology. Let $\mathcal{M}$ be a topological space and consider a map from the circle $\mathbb{S}^1$ to $\mathcal{M}$. We can compose two such maps by merging the two circles, as shown in fig. 1.26. Two maps are said to be *homotopic* if they can be smoothly deformed into each other. Any two homotopic maps are said to belong to the same *equivalence class* or *homotopy class*. For general $\mathcal{M}$, the homotopy classes may be multiplied using the composition law, resulting in a group structure. The group is called the *fundamental group* of the manifold $\mathcal{M}$, and is abbreviated $\pi_1(\mathcal{M})$. If $\mathcal{M} = \mathbb{S}^2$, then any such map can be smoothly contracted to a point on the 2-sphere, which is to say a trivial map. We then have $\pi_1(\mathcal{M}) = 0$. If $\mathcal{M} = \mathbb{S}^1$, the maps can wind nontrivially, and the homotopy classes are labeled by a single integer winding number: $\pi_1(\mathbb{S}^1) = \mathbb{Z}$. The winding number of the composition of two such maps is the sum of their individual winding numbers. If $\mathcal{M} = \mathbb{T}^2$, the maps can wind nontrivially around either of the two cycles of the 2-torus. We then have $\pi_1(\mathbb{T}^2) = \mathbb{Z}^2$, and in general $\pi_1(\mathbb{T}^n) = \mathbb{Z}^n$. This makes good sense, since an $n$-torus is topologically equivalent to

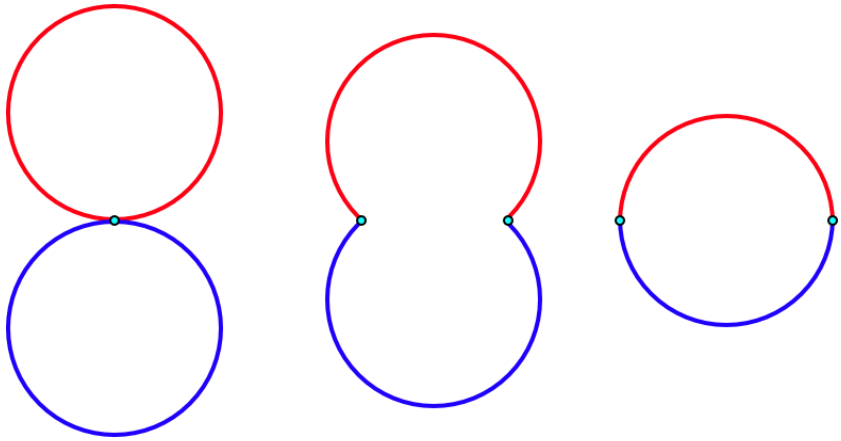

Figure 1.26: Composition of two circles. The same general construction applies to the merging of $n$-spheres $\mathbb{S}^n$, called the *wedge sum*.

a product of $n$ circles. In some cases, $\pi_1(\mathcal{M})$ can be nonabelian, as is the case when $\mathcal{M}$ is the genus $g = 2$ structure shown in the right hand panel of fig. 1.25.

In general we define the $n^{\text{th}}$ *homotopy group* $\pi_n(\mathcal{M})$ as the group under composition of maps from $\mathbb{S}^n$ to $\mathcal{M}$. For $n \geq 2$, $\pi_n(\mathcal{M})$ is abelian. If $\dim(\mathcal{M}) < n$, then $\pi_n(\mathcal{M}) = 0$. In general, $\pi_n(\mathbb{S}^n) = \mathbb{Z}$. These $n^{\text{th}}$ homotopy classes of the $n$-sphere are labeled by their associated Pontrjagin index $Q$.