Title
Topics in topological band systems /

Permalink
https://escholarship.org/uc/item/2vp480dv

Author
Huang, Zhoushen

Publication Date
2014

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA, SAN DIEGO

Topics in topological band systems

A dissertation submitted in partial satisfaction of the requirements for the degree
Doctor of Philosophy

in

Physics

by

Zhoushen Huang

Committee in charge:

Professor Daniel P. Arovas, Chair
Professor Jorge Hirsch
Professor David Meyer
Professor Jeffrey M. Rabin
Professor Congjun Wu

2014
The dissertation of Zhoushen Huang is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

________________________________________

________________________________________

________________________________________

________________________________________

________________________________________

Chair

University of California, San Diego

2014
DEDICATION

To my parents,
for their love and endurance
# TABLE OF CONTENTS

Signature Page ................................................................. iii

Dedication ................................................................. iv

Table of Contents ........................................................... v

List of Figures ............................................................ ix

List of Tables ............................................................... xi

Acknowledgements ........................................................... xii

Vita .................................................................. xiii

Abstract of the Dissertation .................................................. xiv

Chapter 1 Introduction ......................................................... 1

Chapter 2 Boundary states ...................................................... 7
  2.1 Boundary spectrum via $T$-matrix .................................... 7
  2.1.1 Formulation ............................................................ 7
  2.1.2 Zigzag edge states in Graphene ................................. 10
  2.1.3 Zigzag edge states in Haldane model ......................... 11
  2.1.4 Hofstadter model ................................................... 13
  2.2 Zigzag edge state solution in Haldane’s honeycomb model,
  and other related models .................................................. 13
  2.2.1 Haldane’s honeycomb model ................................... 15
  2.2.2 Twisted-boundary Hamiltonian and gauge transforma-
        tion ................................................................ 17
  2.2.3 Edge state Ansatz and energy ................................... 19
  2.2.4 Edge modes crossing points ..................................... 22
  2.2.5 Relation to 1D topological insulators ......................... 24
  2.2.6 Application to edge/end states of other models .......... 25

Chapter 3 Quantum entanglement ........................................... 26
  3.1 General formulation .................................................... 27
  3.2 Real-space entanglement of a single particle wavefunction 28
    3.2.1 Self entanglement of a single particle ..................... 28
    3.2.2 Entanglement of a Bose condensate ....................... 31
  3.3 Real-space entanglement of free fermions ......................... 34
    3.3.1 Relation between a Gaussian density matrix and
          its one-body correlation matrix ......................... 34
Appendix A  Zigzag edge states of the Haldane model, and their application
to graphene and boron nitride ........................................ 170
A.1  Edge solution ......................................................... 170
A.2  Edge crossing point .................................................. 171
A.3  Eigenstates ............................................................ 172
A.4  Graphene and boron nitride ......................................... 175

Appendix B  Incommensurate edge spectrum in the Hofstadter model ... 178
B.1  Review of commensurate edge spectrum ........................... 178
B.2  Incommensurate edge spectrum .................................... 180

Appendix C  Triangular lattice Hofstadter model ....................... 183

Appendix D  Green’s functions of discrete and continuum Weyl Semimetal
models ................................................................. 188
D.1  Spectrum and Green’s function of Eq. 7.2 ......................... 188
D.2  Green’s function of the BHB theory with prototypical \( \tilde{\Gamma} \) . 191
    D.2.1  WS with point nodes (\( \tilde{\Gamma} = \Gamma^{21} \)) .................. 192
    D.2.2  WS with nodal line (\( \tilde{\Gamma} = \Gamma^{35} \)) .................. 195

Bibliography ............................................................. 199
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 2.1</td>
<td>Edge spectrum of Haldane model using $T$-matrix</td>
<td>12</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>Edge spectrum of Hofstadter model using $T$-matrix</td>
<td>14</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>Schematics of Haldane’s honeycomb model</td>
<td>14</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>3D Contour of the $B$ vector in Haldane model</td>
<td>16</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>Zigzag edge modes of Haldane model</td>
<td>21</td>
</tr>
<tr>
<td>Figure 5.1</td>
<td>Energy and entanglement spectrum of the Hofstadter model with $p/q = 3/7$ and three filled bands</td>
<td>48</td>
</tr>
<tr>
<td>Figure 5.2</td>
<td>Energy and entanglement spectrum of the Hofstadter model with $p/q = 3/7$ and two filled bands</td>
<td>49</td>
</tr>
<tr>
<td>Figure 5.3</td>
<td>Full system band projectors for $p/q = 3/7$, $N_y = 28$, with periodic boundary conditions in $y$ and $k_x = 2\pi\kappa/q$.</td>
<td>55</td>
</tr>
<tr>
<td>Figure 5.4</td>
<td>$(m,\kappa)$ construction for $p/q = 3/7$</td>
<td>59</td>
</tr>
<tr>
<td>Figure 5.5</td>
<td>Wannier centers vs. entanglement occupancy</td>
<td>67</td>
</tr>
<tr>
<td>Figure 5.6</td>
<td>Normalized eigenfunctions of $\mathcal{GR}$.</td>
<td>68</td>
</tr>
<tr>
<td>Figure 5.7</td>
<td>Normalized eigenfunctions of $\mathcal{RG}$</td>
<td>70</td>
</tr>
<tr>
<td>Figure 6.1</td>
<td>Uhlmann spectral flow of the Haldane model</td>
<td>89</td>
</tr>
<tr>
<td>Figure 6.2</td>
<td>$\text{Tr} M$ at $k_x = k_c$ for the Haldane model in the parameter space of temperature $T$ and chemical potential $\mu$</td>
<td>92</td>
</tr>
<tr>
<td>Figure 6.3</td>
<td>Uhlmann spectral flow of the BHZ model</td>
<td>93</td>
</tr>
<tr>
<td>Figure 6.4</td>
<td>Uhlmann spectral flow of the Hofstadter model</td>
<td>95</td>
</tr>
<tr>
<td>Figure 7.1</td>
<td>Various ways to identify bound states and resonances for a scalar impurity $V = g\mathcal{I}$ in graphene</td>
<td>103</td>
</tr>
<tr>
<td>Figure 7.2</td>
<td>Resonance condition for $\mathcal{T}$ and $\mathcal{I}$-symmetric Dirac Semimetals</td>
<td>114</td>
</tr>
<tr>
<td>Figure 7.3</td>
<td>Stability of Weyl semimetal with $\mathcal{I}$ breaking $\tilde{\Gamma} = \Gamma^{14}$.</td>
<td>118</td>
</tr>
<tr>
<td>Figure 7.4</td>
<td>Stability of Weyl semimetal with $\mathcal{I}$-symmetric and $\mathcal{T}$ breaking $\tilde{\Gamma} = \Gamma^{12}$.</td>
<td>122</td>
</tr>
<tr>
<td>Figure 7.5</td>
<td>Effect of different classes of impurity ensembles on the ATA DOS.</td>
<td>124</td>
</tr>
<tr>
<td>Figure 8.1</td>
<td>Kitaev’s honeycomb model</td>
<td>128</td>
</tr>
<tr>
<td>Figure 8.2</td>
<td>1D spin lattice</td>
<td>136</td>
</tr>
<tr>
<td>Figure 8.3</td>
<td>1D spin ladder with off-diagonal links</td>
<td>139</td>
</tr>
<tr>
<td>Figure 8.4</td>
<td>Two interpretations of a Majorana spectrum</td>
<td>140</td>
</tr>
<tr>
<td>Figure 8.5</td>
<td>Minimal unit cell bond configuration on the 2D square-octagon lattice</td>
<td>141</td>
</tr>
<tr>
<td>Figure 8.6</td>
<td>Octahedron and its planar representation</td>
<td>145</td>
</tr>
<tr>
<td>Figure 8.7</td>
<td>Octahedron flux classification</td>
<td>146</td>
</tr>
<tr>
<td>Figure 8.8</td>
<td>Flux to link mapping for a single octagon</td>
<td>147</td>
</tr>
<tr>
<td>Figure 8.9</td>
<td>MUC ground state energy of different flux classes relative to class $A$</td>
<td>148</td>
</tr>
</tbody>
</table>
LIST OF TABLES

| Table 7.1: | Symmetry of $\Gamma$ matrices. | 107 |
| Table 7.2: | Impurity classification for $I$ breaking Weyl material. | 116 |
| Table 7.3: | Impurity classification for $I$-symmetric (hence $T$ breaking) Weyl material. | 120 |
| Table 8.1: | Symmetry properties of the square-octagon Hamiltonian | 142 |
| Table 8.2: | Offdiagonal blocks of the OCL Hamiltonian in the octahedron limit. | 148 |
ACKNOWLEDGEMENTS

I am grateful to my advisor, Professor Arovas, for his guidance and unreserved sharing of knowledge. I appreciate the company of my friends at UCSD, without whom life would be less interesting. I thank my thesis committee members for their advices and comments, especially Professor Rabin, for his detailed feedback after reading a draft of this thesis. I thank Professor Balatsky for his hospitality during my visit to LANL.

VITA

2014 Ph.D. in Physics, University of California, San Diego, USA
2007 M.Phil. in Physics, Hong Kong University of Science and Technology, Hong Kong
2004 B.Sc. in Physics, Nanjing University, China

PUBLICATIONS


Da Wang, Zhoushen Huang, and Congjun Wu, “Fate and remnants of Majorana zero modes in a quantum wire array”, Physical Review B 89, 174510, 2014


ABSTRACT OF THE DISSERTATION

Topics in topological band systems

by

Zhoushen Huang

Doctor of Philosophy in Physics

University of California, San Diego, 2014

Professor Daniel P. Arovas, Chair

The discovery of integer quantum Hall effect and its subsequent theoretical formulation heralded a new paradigm of thinking in condensed matter physics, which has by now blossomed into the rapidly growing field of topological phases. In this work we investigate several mutually related topics in the framework of topological band theory. In Chapter 2, we study solutions to boundary states on a lattice and see how they are related to the bulk topology. To elicit a real space manifestation of the non-trivial topology, the presence of a physical edge is not strictly necessary. We study two other possibilities, namely the entanglement spectrum associated with an imaginary spatial boundary, and the localization centers of Wannier functions, in Chapters 3, 4, and 5. Topological classification through discrete indices is so far possible only for systems described by pure quantum states –
in the existing scheme, quantization is lost for systems in mixed states. In Chapter 6, we present a program through which discrete topological indices can be defined for topological band systems at finite temperature, based on Uhlmann’s parallel transport of density matrices. The potential of topological insulators in realistic applications lies in the existence of Dirac nodes on its surface spectrum. Dirac physics, however, is not exclusive to TI surfaces. In a recently discovered class of materials known as Weyl semimetals, energy nodes which emit linear dispersions also occur in the bulk material. In Chapter 7, we study the possibility of resonance states induced by localized impurities near the nodal energy in Weyl semimetals, which will help us in understanding the stability of density-of-state suppression at the energy nodes. Finally, in Chapter 8, we apply the topological characterization developed for noninteracting particles to a class of interacting spin models in 3D, which are generalizations of Kitaev’s honeycomb model, and identify several exotic quantum phases such as spin metals and spin semimetals.
Chapter 1

Introduction

The discovery of the integer quantum Hall effect and its subsequent theoretical formulation heralded a new paradigm of thinking in condensed matter physics, which has by now blossomed into the rapidly growing field of topological phases [26, 54]. In integer quantum Hall systems, the Hall conductance $\sigma_{xy}$ is quantized in units of $e^2/h$, to an astonishing precision of 1 part in $10^9$. Such a quantization is mathematically attributed to a finite integer – the first Chern number $C$ – which can be extracted from the single particle wave functions of the filled bands of the system, as first pointed out in a seminal paper by TKNN [64]. For a single band $|\Psi(k)\rangle$, the Chern number is the integrated Berry curvature over the first Brillouin Zone in units of $2\pi$: $C = \frac{1}{2\pi} \int_{BZ} d^2k \mathcal{F}(k)$ where $\mathcal{F}(k) = \hat{z} \cdot \nabla \times A(k)$ is the Berry curvature and $A(k) = i \langle \Psi(k)|\nabla\Psi(k)\rangle$ is the Berry connection. Here the gradient $\nabla$ is taken in $k$-space. One can think of $A(k)$ as a vector potential in $k$ space, analogous to the electromagnetic vector potential in coordinate space. Then $\mathcal{F}$ plays the role of a “magnetic field” emanating from a “magnetic monopole” the “charge” of which is $C$. For multiple filled bands, $C$ is the sum of Chern numbers of the participating bands. Since an integer cannot change continuously, $\sigma_{xy}$ is robust against perturbations to the system as long as the bulk energy gap remains open, and is said to be topologically protected.

The notion of Chern number underlies the classification of other topological systems as well. In two-dimensional (2D) symmetry-protected topological (SPT) systems, the Chern number itself may vanish per symmetry requirement, for ex-
ample, it must be zero in time reversal (TR)-invariant systems because the Hall conductance is TR odd. However, one can still define a Chern number using a restricted set of wave functions—e.g. a subset of bands \([59, 50]\), or within a subspace in the Brillouin zone \([22]\), etc.—which remains a nonzero integer, an example being the spin Chern number in quantum spin Hall systems. The parity of the Chern number thus defined serves as a topological index for \(Z_2\) topological insulators. For interacting systems whose ground state must be described by an \(N\)-body wavefunction \(\Psi(r_1, r_2, \cdots, r_N)\), one can employ a twisted boundary condition on the \(n\)th (say) coordinate, \(\Psi(r_1, \cdots, r_{n-1}, r_n + L_i, r_{n+1}, \cdots) = e^{i\theta_i} \Psi(r_1, \cdots, r_{n-1}, r_n, r_{n+1}, \cdots)\) where \(L_i\) is the system size in the \(i\)th direction and \(\theta_i\) is the corresponding twist phase. The twist phases replace the Bloch momenta as the adiabatic variables with respect to which a Berry curvature and hence a Chern number can be defined \([46]\). Similar considerations have been generalized to higher spatial dimensions \((D)\) and systems with different global symmetries \((S)\), culminating in classification tables listing all possible bulk topological indices for any combination of \(D\) and \(S\) \([57, 56]\).

The physical significance of the various topological indices lies in what is known as the bulk-boundary correspondence. For integer quantum Hall systems, Hatsugai first showed \([27]\) that on a cylindrical geometry, the net number of edge states flowing across a bulk spectral gap is given by the total Chern number of the bulk bands below the gap. These edge modes are chiral in the sense that they propagate in opposite directions at opposite physical edges. Each edge state can be thought of as a 1D conducting channel with conductance \(\pm e^2/h\) where the sign is determined by their direction of propagation, the overall conductance is thus proportional to the net number of channels, namely the Chern number. In 3D, the boundary consists of surfaces. For strong 3D topological insulators, each surface can host a single Dirac point in its surface energy spectrum, providing a way out of the famous Fermion doubling theorem \([45]\).

Topological characterization not only applies to bulk insulators, but also to bulk semimetals. An ordinary metal has a Fermi surface (FS) with codimension 1 due to the presence of a single constraint \(E(k) = E_F\) where \(E_F\) is the Fermi
energy. Semimetals have a higher FS codimension. For example, in graphene, 
\[ H_{xy}(k) = b_x(k)\sigma_x + b_y(k)\sigma_y, \]
thus the single constraint \( E = 0 \) would entail two independent constraints \( b_x = 0 \) and \( b_y = 0 \), leading to a FS with codimension 2. On the other hand, the existence of FS at all requires the spatial dimension \( D \) to be no less than the FS codimension. For example, an arbitrary two band model would have 
\[ H(k) = H_{xy}(k) + b_z(k)\sigma_z, \]
hence the FS codimension is 3 with the extra condition \( b_z = 0 \). This over-constrains the \( D = 2 \) momentum space, thus two band models in 2D are generically gapped. To salvage FS, one appeals to symmetry. In graphene, the sublattice symmetry guarantees \( b_z(k) = 0 \ \forall k \), eliminating the \( \sigma_z \) term. In general, global symmetries force a subset of the FS constraints to be identically zero, in other words, the energy nodes in semimetals require symmetry protection. Once protected, these nodes are subject to topological classifications. For example, a point node in 2D semimetals can be classified by the Berry phase accumulated over a loop in \( k \)-space surrounding it; in 3D, one can compute a Chern number over a \( k \)-space surface surrounding the node. Bulk-boundary correspondence then demands that edge/surface modes connect between the projection of these nodes onto the edge/surface Brillouin zone. The famous zigzag edge zero mode in graphene is a direct consequence of the Dirac nodes.

In this work we will investigate several mutually related topics in the framework of topological band theory. In Chapter 2, we will study some solutions for boundary states on a lattice and see how they are related to the bulk topology, a manifestation of bulk-boundary correspondence. The presence of a boundary can be treated as a special kind of defect in a periodic system, thus its effect can in general be studied by the \( T \)-matrix formalism standard in solving defect problems. Such a treatment is general but less transparent than a direct solution of the eigenvalue problem of the Hamiltonian. For this purpose we also present a direct solution for certain types of Hamiltonians, and illustrate it in detail by solving the zigzag edge state in Haldane’s honeycomb model. While this latter solution relies on the form of the Hamiltonian and thus cannot be generic, it does cover the end/edge states of several other interesting and paradigmatic models such as graphene, Boron-Nitride, Creutz ladder [18], Su-Schrieffer-Heeger [63], Kitaev
The intriguing aspect of the bulk-boundary correspondence, in some sense, is in the connection it establishes between two complementary viewpoints, namely momentum space and real space. A natural question to ask is if the presence of a physical edge is necessary to elicit a real space manifestation. The answer is No, and there are two routes of circumvention. The first is to replace the physical boundary with an “imaginary” one, and we shall explore this line in Chapters 3 and 5 by studying the entanglement spectrum with respect to a real space bipartition. In the entanglement problem, one conceptually divides a complete system into two parts, then each part needs to be described by a density matrix instead of a pure state, and their interface, i.e. the entanglement cut, serves as the imaginary boundary. Spatial directions parallel to the entanglement cut can still be treated as translationally invariant, thus one can organize the entanglement spectrum as functions of the transverse momenta $k_\perp$ much like the energy spectrum in the presence of a physical boundary. As we shall see, in topologically nontrivial systems, the entanglement spectrum exhibits spectral flow, with respect to $k_\perp$, analogous to that of the edge state energies, although the entanglement cut is a purely conceptual device. The second line of investigation eliminates real space boundaries altogether, physical or imaginary, and instead proceeds by recombining momentum eigenstates such that the resulting states are localized in one spatial direction – essentially a partial inverse Fourier transform. We will study these recombined states, known as hybrid Wannier states, in Chapters 4 and 5. As we shall see, their localization centers in the chosen direction are offset from unit cell boundaries by an amount of the Berry phase (in unit of $2\pi$) of the band to which they belong. Like the edge energies and entanglement spectra, the Wannier centers also exhibit spectral flow, but without invoking any notion of a boundary. We shall establish a one-to-one mapping between entanglement and Wannier spectrum, providing a unifying framework for both phenomena.

As discussed before, the integer characterization of topological insulators relies on the Chern number or its variants, which have no simple generalization to finite temperature. This is because thermal average of the Chern number would
include a contribution from bands not filled at zero temperature, and would therefore cease to be an integer. In Chapter 6, we will present a program through which one can define topological indices of TIs at finite temperature. The idea is the following: Recall that the Chern number of a 2D system can be formulated as a winding number with respect to one Bloch momentum, say $k_x$, of its Berry phase computed along $k_y = 0 \to 2\pi$. While the Berry phase can only be defined for pure states and therefore cannot be used for open systems, one can nevertheless define one or more geometric phases for a cyclic path of density matrices. Once such geometric phases are identified, their winding numbers can be used as a topological characterization for open systems. In Chapter 6, we will adopt Uhlmann’s parallel transport for density matrices [66] to define the geometric phases of open systems, and demonstrate the scheme described above in model systems.

The appeal of 3D topological insulators to possible practical applications lies in their surfaces, where the low energy physics can be described by the Dirac equation. The Dirac-type physics, however, is not exclusive to TI surfaces. Recently a new type of material known as Weyl semimetals has attracted much attention due to its possible realization in real materials [70, 10]. In Weyl semimetals, the bulk itself contains energy nodes with linear dispersion. The name “Weyl” comes from the fact that near the energy nodes, the physics can be formulated in terms of 2-component Weyl fermions. As discussed before, these nodes can be classified by the Chern number computed over a $k$-space surface surrounding them, and through bulk-boundary correspondence, imply the existence of Fermi arcs, i.e. the locus of gapless surface states interpolating between the projections of Weyl nodes on the surface Brillouin zone. In Chapter 7, we study the possibility of inducing resonant states near the Fermi surface through various forms of localized impurity. The purpose is two-fold: (1) As discussed before, the hallmark of semimetals is in the suppression of the density of states (DOS) at the nodal energy. Since such a suppression underlies various proposals for its practical application, and since impurities are always present in real materials, one should ask if it can be lifted due to impurity effects. (2) We have mentioned that a physical boundary can be viewed as a special type of defect, and in topological materials, such “defects” can
trap low energy in-gap states through bulk-boundary correspondence. The question is if this capability extends to localized impurities, which can be viewed as a miniature version of physical boundaries. These questions will be addressed in Chapter 7.

The topological characterization developed for noninteracting particles also finds application in interacting systems. In a seminal paper [38], Kitaev proposed a model of interacting spin-$\frac{1}{2}$s on a honeycomb lattice, which can be solved exactly by a mapping of the spins to free Majorana fermions hopping in static $Z_2$ gauge fields. To find the ground state, for example, one diagonalizes the quadratic Majorana Hamiltonian and finds the lowest energy state for each $Z_2$ configuration. The true ground state is then the global minimum of these lowest energy states. In fact the energy of any desired flux sector could be lowered arbitrarily by adding certain types of interaction terms to the Hamiltonian which do not affect solvability, thus being a true ground state is no longer a very restrictive requirement. In Chapter 8, we will first clarify certain issues related to the Majorana nature of the solution in 1D and 2D, and then generalize Kitaev’s model to two 3D lattices, the octagon-cubic lattice and the pyrochlore lattice, in which we identify several exotic quantum phases such as spin metals and spin semimetals, to borrow terms from band insulators.
Chapter 2

Boundary states

Hatsugai [27] first pointed out that in 2D Chern insulators, the physical significance of nontrivial bulk topology is that in the presence of physical edges, the edge states exhibit energy spectral flows which interpolate the bulk gap, thereby providing conduction channels responsible for the quantum Hall effect. States at opposite edges propagate in opposite directions, viz. with opposite velocity $d\varepsilon_k/dk$, hence they must become degenerate at some momentum $k$, which normally would be avoided due to Wigner-von Neumann Theorem. This is the most basic form of bulk-boundary correspondence. In 1D, nontrivial topology manifests as degenerate in-gap end states, whereas in 3D it appears as surface Dirac cones. In this chapter we study two approaches to solving boundary states on a lattice.

2.1 Boundary spectrum via $T$-matrix

2.1.1 Formulation

A tight-binding Hamiltonian with periodic boundary conditions (PBC) is easily diagonalized using Bloch’s Theorem. The presence of a physical boundary—surface for 3D, edge for 2D, and end for 1D—is modeled by eliminating matrix elements across the boundary, resulting in a Hamiltonian with the so-called open boundary condition (OBC). We shall focus on 1D. For higher dimensions, we assume OBC in only one direction and PBC in the rest, then transverse momenta
(those parallel to the boundary) are good quantum numbers and can be treated as parameters of an effectively 1D Hamiltonian.

Consider a 1D translationally invariant Hamiltonian $\hat{H} = m \sum_i c_i^\dagger c_i + t \sum_{\langle ij \rangle} c_i^\dagger c_j \equiv \sum_{ij} H_{ij} c_i^\dagger c_j$ on a lattice of $N$ sites, where $i, j = 1, 2, \cdots, N$ label the sites and $\langle \cdots \rangle$ denotes nearest neighbors. The effect of boundary condition is that for PBC, there is a hopping term between sites 1 and $N$, whereas for OBC this term is absent. In terms of the coefficient matrix $H_{ij}$, one has

$$H = \begin{pmatrix} m & t & z \\ t & m & t \\ \vdots & \ddots & \ddots \\ z & t & m \end{pmatrix}, \quad z = \begin{cases} t & \text{PBC} \\ 0 & \text{OBC} \end{cases}. \quad (2.1)$$

Thus one may write

$$H_{OBC} = H_{PBC} - t \left[ |1\rangle \langle N| + |N\rangle \langle 1| \right]. \quad (2.2)$$

The idea then is to treat the terms in bracket as a localized perturbation to $H_{PBC}$.

In general, the Hamiltonian is translationally invariant over the scale of a unit cell instead of a single site. A unit cell may include various degrees of freedom such as sublattices, orbitals, and spins, the total number of which we shall denote as $Q$. Mathematically, one replaces each matrix element of Eq. 2.1 with a $Q \times Q$ block accounting for the internal structure of a unit cell, viz. $m \to h$, $t \to v$ and $t^* \to v^\dagger$, yielding

$$H_{PBC} = \mathbb{I}_N \otimes h + \left[ \mathbb{D} \otimes v + h.c. \right], \quad (2.3)$$

where $\mathbb{D}$ is a superdiagonal matrix of 1 implementing a translation of the PBC lattice by one unit cell,

$$\mathbb{D} = \begin{pmatrix} \mathbb{I}_{N-1} \\ 1 \end{pmatrix}, \quad (2.4)$$

and the corresponding OBC Hamiltonian is

$$H_{OBC} = H_{PBC} + V, \quad (2.5)$$
where $V$ encodes the effect of the OBC,

$$V = -|1\rangle\langle N| \otimes v^\dagger - |N\rangle\langle 1| \otimes v$$

$$= \begin{pmatrix} v & v^\dagger \end{pmatrix}. \quad (2.6)$$

In the second line we introduced a block matrix representation of $V$ in the subspace of $\{|1\rangle, |N\rangle\}$. End states of $H_{\text{OBC}}$ are bound states of $V$ and can be obtained from the $T$-matrix approach [19].

A more detailed account of the application of the $T$-matrix in multi-band systems will be given in Chapter 7. Here let us just recall that for a generic Hamiltonian $H = H^{(0)} + V$, its Green’s function is $G(\omega) = (\omega - H)^{-1} = G^{(0)}(\omega) + G^{(0)}(\omega)T(\omega)G^{(0)}(\omega)$, where the $T$ matrix associated with $V$ is $T(\omega) = V(\mathbb{I} - G^{(0)}(\omega)V)^{-1} = V + VG^{(0)}(\omega)V + VG^{(0)}(\omega)VG^{(0)}(\omega)V + \cdots$ and $G^{(0)}(\omega) = (\omega - H^{(0)})^{-1}$ is the Green’s function of $H^{(0)}$. By construction, $G(\omega)$ diverges if $\omega$ is an eigenvalue of $H$, thus a state induced by $V$ with energy $E = \omega$ will cause $T(\omega)$ to diverge.

For a generic $G^{(0)}$, $T$ has the same rank as $V$ and can be inverted in its column subspace, i.e. the subspace orthogonal to the kernel of $V$,

$$T^{-1}(\omega) = V^{-1} - P_V G^{(0)}(\omega) P_V \quad (2.7)$$

where $P_V$ is the projector onto the column subspace of $V$. Taking $H^{(0)} = H_{\text{PBC}}$ and $V$ as in Eq. 2.6, $P_V$ is simply $|1\rangle\langle 1| + |N\rangle\langle N|$ when $v$ is invertible\(^1\), and analogous to the second line of Eq. 2.6 we may write

$$P_V G^{(0)}(\omega) P_V = \begin{pmatrix} G_{11}^{(0)}(\omega) & G_{1N}^{(0)}(\omega) \\ G_{N1}^{(0)}(\omega) & G_{NN}^{(0)}(\omega) \end{pmatrix} = \left\langle \begin{pmatrix} G^{(0)}(k,\omega) & G^{(0)}(k,\omega)e^{ik} \\ G^{(0)}(k,\omega)e^{-ik} & G^{(0)}(k,\omega) \end{pmatrix} \right\rangle_k$$

where $G^{(0)}(k,\omega) = (\omega - H^{(0)}(k))^{-1}$ is the $k$-space Green’s function of $H^{(0)}$, with $H^{(0)}(k) = h + ve^{ik} + v^\dagger e^{-ik}$, and $\langle \cdots \rangle_k$ denoting averaging over $k$. The end states are energies $E_B$ such that $T(E_B)$ diverges, or equivalently, $T^{-1}(E_B)$ has a zero

\(^1\)See §2.1.2 for a treatment where $v$ is not invertible.
mode. For $E_B$ in the bulk bands, one may replace $G^{(0)}(k, E_B)$ with its hermitian part $G^{(0)}(k, E_B) \equiv \left[ G^{(0)}(k, E_B + i\epsilon) + (G^{(0)}(k, E_B + i\epsilon))^\dagger \right]/2$ with $\epsilon \to 0^+$, see Chapter 7.

Let us now look at some examples.

### 2.1.2 Zigzag edge states in Graphene

The Hamiltonian of graphene with zigzag edge can be written as

$$ h(k_1) = \begin{pmatrix} 0 & 1 + e^{-ik_1} \\ 1 + e^{ik_1} & 0 \end{pmatrix}, \quad v = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \quad (2.9) $$

where $k_1 = k \cdot a_1$ is the momentum along the zigzag edge ($a_1$ direction). It is well known that there are two degenerate edge states at $E = 0$ for $k_1 \in \left( \frac{2\pi}{3}, \frac{4\pi}{3} \right)$. We shall verify this using the $T$-matrix formalism and show its connection with a $\pi$ Berry phase.

The PBC Hamiltonian is

$$ H^{(0)}(k) = h(k_1) + ve^{ik_2} + v^*e^{-ik_2} = \begin{pmatrix} 0 & \gamma^*(k) \\ \gamma(k) & 0 \end{pmatrix} \quad (2.10) $$

where $\gamma(k) \equiv 1 + e^{ik_1} + e^{ik_2}$ and $k_2 = k \cdot a_2$ is the momentum along the $a_2$ basis vector. The $k$-space Green’s function is

$$ G^{(0)}(k, \omega) = \frac{\omega + H^{(0)}(k)}{\omega^2 - |\gamma(k)|^2}. \quad (2.11) $$

Note that $v$ is not invertible. In the edge subspace,

$$ V = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \implies P_V = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \quad (2.12) $$

The nonvanishing block of $P_VG^{(0)}P_V$ is thus a $2 \times 2$ matrix,

$$ P_VG^{(0)}(k_1, \omega)P_V \equiv \begin{pmatrix} G^{(0)}_{11}(k, \omega) & G^{(0)}_{12}(k, \omega)e^{ik_2} \\ G^{(0)}_{21}(k, \omega)e^{-ik_2} & G^{(0)}_{22}(k, \omega) \end{pmatrix}_{k_2}. \quad (2.13) $$
\langle \cdots \rangle_{k_2} \text{ denotes averaging over } k_2.

Let us now investigate the possibility of zero energy modes induced by \( V \). Restricting to \( \omega = 0 \) (for zero energy), we have \( V^{-1} = -\sigma_x \) and \( G^{(0)}(k) = -H^{(0)}(k)/|\gamma(k)|^2 = - \begin{pmatrix} 0 & 1/\gamma(k) \\ 1/\gamma^*(k) & 0 \end{pmatrix} \), thus

\[
T^{-1}(k_1, \omega = 0) = \begin{pmatrix} 0 & g_{12}(k_1) - 1 \\ g_{12}^*(k_1) - 1 & 0 \end{pmatrix}
\]  

(2.14)

with

\[
g_{12}(k_1) = -\langle G_{12}^{(0)}(k, \omega = 0)e^{ik_2}\rangle_{k_2} = \int_0^{2\pi} \frac{dk_2}{2\pi} \frac{e^{ik_2}}{\gamma(k)}
\]

\[
= \oint_{|z|=1} \frac{dz}{2\pi i z + q(k_1)} = \begin{cases} 
1 & \text{if } |q(k_1)| < 1 \\
0 & \text{if } |q(k_1)| > 1
\end{cases}
\]

(2.15)

We have defined \( z \equiv e^{ik_2} \) and \( q \equiv 1 + e^{ik_1} \). \( T^{-1} \) is identically zero if \( q(k_1) \) is outside the unit circle, viz. \( k_1 \in \left( \frac{2\pi}{3}, \frac{4\pi}{3} \right) \). This verifies that \( E = 0 \) zigzag edge states exist in this region – and since both eigenvalues of \( T^{-1} \) are zero, there should be two \( E = 0 \) states (localized at opposite edges).

Note that as \( k_2 \) sweeps a \( 2\pi \) cycle, \( \gamma(k) \) in Eq. 2.10 may or may not wind in the complex plane depending on the value of \( k_1 \). If one were to compute the Berry phase of, say, the lower band of Eq. 2.10 over the \( k_2 \) cycle, it would be \( \pi \) if \( \gamma \) winds once, and 0 if \( \gamma \) does not wind. Since the winding number is given exactly by Eq. 2.15, we see that a \( \pi \) Berry phase guarantees the existence of degenerate edge modes.

### 2.1.3 Zigzag edge states in Haldane model

Haldane’s honeycomb model will be discussed in more detail in §2.2. Here we only write down its Hamiltonian with zigzag edge,

\[
h(k_1) = \begin{pmatrix} h_1 & p_1 \\ p_1^* & h_2 \end{pmatrix}, \quad v(k_1) = \begin{pmatrix} v_1 & 0 \\ p_2 & v_2 \end{pmatrix}
\]

(2.16)
Figure 2.1: Edge spectrum of Haldane model using T-matrix. (a) The logarithm of the operator norm of the boundary T matrix, $||T(E, k_1)||$ diverges if $E(k_1)$ is an edge energy. (b) The correction to the LDOS at the $r_2 = 1$ edge, which should extract the lower edge spectrum. Parameters used: $[t_1, t_2, t_3] = [0.3, 0.4, 0.5]$, $m = 1.4$, $\phi = 0.3\pi$. $N_2 = 50$ unit cells used in the $a_2$ direction, and spectral broadening is $\epsilon = 0.01$.

where

\begin{align*}
h_1 &= m - 2t_1 \cos(\phi + k_1) \quad , \quad h_2 = -m - 2t_1 \cos(\phi - k_1) \quad , \quad p_1 = -1 - e^{-ik_1}, \\
v_1 &= -t_2 e^{-i\phi} - t_3 e^{i\phi} e^{-ik_1} \quad , \quad v_2 = -t_2 e^{i\phi} - t_3 e^{-i\phi} e^{-ik_1} \quad , \quad p_2 = -1. \quad (2.17)
\end{align*}

We solve the edge spectrum numerically, and plot it in Fig. 2.1. In panel (a) states localized on both edges are plotted, in panel (b), only states localized on one edge (the lower edge) are plotted.
2.1.4 Hofstadter model

The Hofstadter model will be discussed in more detail in Chapter 5. Here we only write down its Hamiltonian with open boundary condition,

\[
h(k_x) = -\begin{pmatrix} c_1 & 1 \\ 1 & c_2 & 1 \\ & \ddots & \ddots \\ & & c_{q-1} & 1 \\ & & 1 & c_q \end{pmatrix}, \quad v = -|q\rangle\langle 1|. \tag{2.18}
\]

where

\[
c_a = 2 \cos \left[ k_x + (a - 1) \times 2\pi p/q \right] \tag{2.19}
\]

and the flux per square plaquette is $2\pi p/q$ with co-prime integers $p$ and $q$. Note that $v$ is non-invertible just like in the graphene case. Thus $T^{-1}$ is a $2 \times 2$ matrix.

We solve the edge spectrum numerically and plot it in Fig. 2.2. In panel (a) states localized on both edges are plotted, in panel (b), only states localized on one edge (the lower edge) are plotted.

2.2 Zigzag edge state solution in Haldane’s honeycomb model, and other related models

In this section we will solve the zigzag edge state of Haldane’s honeycomb model analytically. Instead of solving an open boundary problem directly, we look for solutions of a particular form which solves the problem in the bulk but on the edges would require a \textit{generalized twisted boundary condition}, to be defined later. We identify these solutions as edge states when their boundary condition reduces to that of an open edge. Material in this section is based on an earlier manuscript [31].
Figure 2.2: Edge spectrum of Hofstadter model using $T$-matrix. (a) The logarithm of the operator norm of the boundary $T$ matrix. $||T(E,k_x)||$ diverges if $E(k_x)$ is an edge energy. (b) The correction to the LDOS at the $y = 1$ edge, which should extract the lower edge spectrum. Parameters used: $p = 3, q = 7$. $N_2 = 50$ unit cells used in the $\hat{y}$ direction, and spectral broadening is $\epsilon = 0.01$.

Figure 2.3: Haldane’s model. $A$ (red) and $B$ (white) label sublattice sites, and the boxed pair represents a unit cell. Primitive lattice vectors are chosen to be $a_1$ and $a_2$ as shown. Second neighbor hopping between same sublattice sites picks up a phase of $+\phi$ along the arrowed directions and $-\phi$ in opposite directions. Horizontal box encloses a zigzag edge.
2.2.1 Haldane’s honeycomb model

Haldane’s honeycomb model [25] is a tight binding model on the honeycomb lattice, described by the Hamiltonian

\[ H = - \sum_{i,j} t_{ij} c_i^\dagger c_j + \sum_i m_i c_i^\dagger c_i. \] (2.20)

The hopping amplitudes \( t_{ij} \) are nonzero only for nearest neighbor (NN) and next-nearest neighbor (NNN) hopping. For NN hops, \( t_{ij} \equiv t \) is real. For NNN hops, \( t_{ij} = t_s e^{\pm i\phi} \), where \( s = 1 \) if the hopping is parallel to \( a_1 \), \( s = 2 \) if parallel to \( a_2 \), and \( s = 3 \) if parallel to \( a_3 \equiv a_2 - a_1 \). The sign of the phase is according to the arrows in Fig. 2.3, and is taken to be positive for clockwise hops within each hexagonal unit cell. In Haldane’s original model, \( t_1 = t_2 = t_3 \). Setting these amplitudes to be different breaks the three-fold rotational symmetry. The Semenoff mass \( m_i \) is \( m \) and \( -m \) for \( A \) and \( B \) sublattices respectively, which breaks spatial inversion.

In the bulk, the Fourier transformed Hamiltonian is

\[ H(k) = \omega(k) + B(k) \cdot \sigma, \] (2.21)

\( \sigma = (\sigma_x, \sigma_y, \sigma_z) \) are the Pauli matrices in the “isospin” degree of freedom, where \( A \) and \( B \) are isospin up and down, respectively, and

\[
\begin{align*}
B_x &= -1 - \cos k_1 - \cos k_2 \\
B_y &= -\sin k_1 - \sin k_2 \\
B_z &= m + 2[t_1 \sin k_1 - t_2 \sin k_2 + t_3 \sin(k_2 - k_1)] \sin \phi \\
\omega &= -2[t_1 \cos k_1 + t_2 \cos k_2 + t_3 \cos(k_2 - k_1)] \cos \phi
\end{align*}
\] (2.22)

Here, \( k_i = k \cdot a_i \in [0, 2\pi] \) are the Bloch phases along the two primitive vectors. The bulk topology is characterized by the Chern number \( C \) of the upper band, which is the winding number of the unit vector \( \hat{B}(k) \) over the Brillouin zone. That is to say, if by varying \((k_1, k_2)\) over the first Brillouin zone, \( \hat{B} = B/|B| \) covers the unit sphere once \((|C| = 1)\), then the system is in its topological phase, otherwise it is in its trivial phase. Equivalently, in the topological phase the origin is inside the surface swept out by \( B \), while in the nontopological phase it lies outside. The
Figure 2.4: 3D contour of the $B$ vector in Haldane model, see Eq. 2.21. As $k$ varies in the two-dimensional Brillouin zone, $B(k)$ moves in a three-dimensional vector space. The wrapping number of the tip of $B$ around the origin is the Chern number. Greenish mesh lines correspond to loci of $B$ at constant $k_x$ and $k_y$. Blue and red loops are $B(k)$ at $k_x = k_c$, cf. Eq. 2.48. Both loops are coplanar and their respective planes pass through the origin. The blue loop winds around the origin once, the red loop zero time.
topological phase transition thus takes place when the origin is crossed by $B$ at
some $k$ points, where the gap $\Delta = 2|B|$ will vanish. Following Eq. 2.22, this
can only happen at the graphene Dirac points, $(k_1, k_2) = \pm(2\pi/3, -2\pi/3) \equiv K_\pm$, 
where $B_x = B_y = 0$. The corresponding $B_z$ values necessarily have opposite signs
in the topological phase (so that the origin is enclosed), i.e.[25],

$$
|m| - \sqrt{3}|(t_1 + t_2 + t_3)\sin\phi| \begin{cases} < 0 & \text{topological} \\ > 0 & \text{trivial} \end{cases} \quad (2.23)
$$

In the trivial phase, the eigenstates just below the gap at both Dirac points are
entirely concentrated on the same sublattice. In the topological phase, however,
they are concentrated on opposite sublattices.

2.2.2 Twisted-boundary Hamiltonian and gauge transformation

The zigzag edge is parallel to $a_1$ (see the horizontal box in Fig. 2.3), hence
$k_1 = k \cdot a_1$ is a good quantum number. Assume there are $N$ unit cells in the
$a_2$ direction. At each $k_1$, the effective 1-D system is described by the following
Hamiltonian,

$$
\mathcal{H}(k_1) = \sum_{n, n'}^{N} \left[ a_n^\dagger K^{(1)}_{nn'} a_{n'} + b_n^\dagger K^{(2)}_{nn'} b_{n'} + a_{nt} R_{n, nn'} b_{n'} \right. + \left. b_n^\dagger R^\dagger_{n, nn'} a_{n'} \right] \quad (2.24)
$$

Here $a_n^\dagger$ and $b_n^\dagger$ are creation operators on the A and B site of the $n^{th}$ unit cell,
respectively. The coefficient matrices $K^{(i)}$ connect sites on the same sublattice,
and $R$ connects different sublattices. Their nonzero matrix elements are given by

$$
K^{(i)} = \begin{pmatrix}
  h_i & v_i & \tilde{v}_i^* \\
  v_i^* & h_i & v_i \\
  \tilde{v}_i & v_i^* & h_i
\end{pmatrix}, \quad R = \begin{pmatrix}
  p_1 & \tilde{v}_2^* \\
  p_2 & p_1 \\
  \tilde{v}_1 & p_2 & p_1
\end{pmatrix} \quad (2.25)
$$
We will explain the matrix elements \( \bar{v} \) in the corners later. Other matrix elements can be obtained from Fourier transforming the bulk Hamiltonian Eq. 2.21,

\[
h \equiv \sum_{k_2} H(k_1, k_2) = \begin{pmatrix} h_1 & p_1 \\ p_1^* & h_2 \end{pmatrix},
\]

(2.26)

\[
v \equiv \sum_{k_2} H(k_1, k_2) e^{-i k_2} = \begin{pmatrix} v_1 & 0 \\ p_2 & v_2 \end{pmatrix}
\]

(2.27)

with

\[
h_1 = m - 2t_1 \cos(\phi + k_1), \quad (2.28)
\]

\[
h_2 = -m - 2t_1 \cos(\phi - k_1), \quad (2.29)
\]

\[
v_1 = -t_2 e^{-i\phi} - t_3 e^{i\phi} e^{-ik_1}, \quad (2.30)
\]

\[
v_2 = -t_2 e^{i\phi} - t_3 e^{-i\phi} e^{-ik_1}, \quad (2.31)
\]

\[
p_1 = -1 - e^{-ik_1}, \quad p_2 = -1. \quad (2.32)
\]

We note that swapping subscripts 1 and 2 on the left hand sides yields a Hamiltonian with the so-called bearded edge. The method we describe below applies to both types of edge.

The following gauge transformation makes both \( p_1 \) and \( p_2 \) real,

\[
a_n(k_1) \to a_n(k_1) e^{-i(n-1)k_1/2}, \quad b_n(k_1) \to b_n(k_1) e^{-ink_1/2}, \quad (2.33)
\]

by which \( v_1 \to v_1 e^{ik_1/2} \) and \( p_1 \to p_1 e^{ik_1/2}, \) viz.,

\[
v_1 \to -(t_2 + t_3 \cos(\phi - \frac{k_1}{2}) + i(t_2 - t_3) \sin(\phi - \frac{k_1}{2}) \), \quad (2.34)
\]

\[
v_2 \to -(t_2 + t_3 \cos(\phi + \frac{k_1}{2}) - i(t_2 - t_3) \sin(\phi + \frac{k_1}{2}) \),
\]

\[
p_1 \to -2 \cos \frac{k_1}{2}.
\]

In Eq. 2.25, a twisted boundary condition \[53\] is used:

\[
\bar{v} \equiv \rho U v = \begin{pmatrix} \bar{v}_1 & \bar{v}_{12} \\ \bar{v}_{21} & \bar{v}_2 \end{pmatrix}.
\]

(2.35)

\( \rho \) is a real number controlling the “tunnelling strength” between the two edges, and \( U \) is a unitary \( 2 \times 2 \) matrix that describes an “isospin-dependent” phase
twisting over the boundary. For an open boundary, \( \rho \to 0 \). For periodic boundary conditions, \( \rho = 1 \) with \( U = \mathbb{I} \) (without the gauge transformation of Eq. 2.33) or \( U = e^{-iNk_1/2} \mathbb{I} \) (with Eq. 2.33), where \( \mathbb{I} \) is the 2 \( \times \) 2 identity matrix. Introducing twisted boundary condition may seem to overcomplicate the situation, but as we shall see it allows us to make progress toward an analytical solution.

The eigenvalue problem can now be written as

\[
K^{(1)} |\psi_A\rangle + R |\psi_B\rangle = \varepsilon |\psi_A\rangle ,
K^{(2)} |\psi_B\rangle + R^\dagger |\psi_A\rangle = \varepsilon |\psi_B\rangle ,
\]

(2.36)

where \( |\psi_A\rangle \) and \( |\psi_B\rangle \) are the “wavefunctions” of the A and B sublattices, both of which are \( N \)-dimensional column vectors.

### 2.2.3 Edge state Ansatz and energy

We look for solutions of the form \( |\psi_A\rangle = |\psi\rangle \) and \( |\psi_B\rangle = \lambda |\psi\rangle \). In other words, \( |\psi\rangle \) is the real space part and \( (\lambda) \) is the internal space part. Eq. 2.36 now becomes

\[
(K^{(1)} + \lambda R - \varepsilon) |\psi\rangle = 0 ,
(\lambda K^{(2)} + R^\dagger - \lambda \varepsilon) |\psi\rangle = 0 .
\]

(2.37) (2.38)

A sufficient condition for both equations to be satisfied is that the coefficient matrices are proportional element by element,

\[
\frac{h_1 + \lambda p_1 - \varepsilon}{h_2 + \lambda^{-1} p_1 - \varepsilon} = \frac{v_1}{v_2 + \lambda^{-1} p_2} = \frac{v_1^* + \lambda p_2}{v_2^*} \equiv r .
\]

(2.39)

This gives, at each value of \( k_1 \), two equations (the ratio \( r \) itself being yet undetermined) for the two unknowns \( \lambda \) and \( \varepsilon \). The solutions are

\[
\lambda_\pm = \frac{v_1 v_2^* - v_2 v_1^* - p_2^2 \pm \sqrt{d^2 - 4 |v_1 v_2|^2}}{2 p_2 v_2} ,
\]

(2.40)

\[
r_\pm \equiv r(\lambda_\pm) = \frac{d \pm \sqrt{d^2 - 4 |v_1 v_2|^2}}{2 |v_2|^2} ,
\]

(2.41)

\[
\varepsilon_\pm = \frac{r_\pm (p_2 h_2 - 2 p_1 \Re v_2) - (p_2 h_1 - 2 p_1 \Re v_1)}{p_2 (r_\pm - 1)} ,
\]

(2.42)
where $\text{Re}$ indicates the real part, $\pm$ denotes the branch of solution, and

$$d = v_1 v_2^* + v_2 v_1^* - p_2^2 \in \mathbb{R}.$$  \hfill (2.43)

See Appendix A for details regarding derivation.

Eq. 2.41 becomes singular when $v_2 = 0$, which could happen if one simultaneously has $t_2 = t_3$ and $k_1 = \pi - 2\phi$. For this particular parameter set, it can be readily verified, by Taylor expansion in $v_2$, that

$$\lambda_+ = v_1, \quad r_+ = -v_1, \quad \varepsilon_+ = \frac{h_2 v_1^2 + 2p_1 v_1 + h_1}{1 + v_1^2},$$

$$\lambda_-^{-1} = v_2 \to 0, \quad r_-^{-1} = -v_2^2 \to 0, \quad \varepsilon_- = h_2.$$  \hfill (2.44)

These results also hold for the simplified cases of graphene ($m = 0$) and for boron nitride ($m \neq 0$), where second neighbor hoppings are turned off, rendering $v_1 = v_2 = 0$. Clearly, $\lambda_+ = r_+ = \lambda_-^{-1} = r_-^{-1} = 0$, and $\varepsilon_\pm = \pm m$. As will be shown in Appendix A, $\rho \to 0$ if $|k_1| > 2\pi/3$, so solutions there correspond to edge modes with open boundary.

Eq. 2.42 is the central result of this section. Since the reality of $\varepsilon_\pm$ is equivalent to the reality of $r_\pm$, in order to have real energy solutions at a chosen $k_1$, the discriminant of $r_\pm$ must satisfy

$$\Delta = d^2 - 4|v_1 v_2|^2 \geq 0$$  \hfill (2.45)

at the chosen $k_1$.

Although real solutions exist for all $k_1$ satisfying $\Delta(k_1) \geq 0$, they do not necessarily correspond to open boundaries. Normally, wavefunctions are solved after fixing boundary conditions ($\rho$ and $U$), but here, we enforced a particular form of solution, which will not be consistent with arbitrary $\rho U$. Instead, the matrix $\rho U$ is to be determined self-consistently from the Ansatz, and is in general $k_1$-dependent. This is discussed in detail in Appendix A. For now, we just note that only when $\rho \to 0$ will the solution be valid for open boundary. Clearly, as $\rho$ varies with $k_1$, the transition from an open boundary solution ($\rho \to 0$) to that of an “enhanced tunnelling” boundary ($\rho > 1$) will happen when $\rho = 1$, at which point
Figure 2.5: Zigzag edge modes of Haldane model with parameters $[t_1, t_2, t_3] = [0.3, 0.4, 0.5]$, $m = 1.4$, $\phi = 0.3\pi$, $N = 40$. (a) Circular dots are obtained from diagonalizing an open boundary Hamiltonian, Eq. 2.24 with $\rho = 0$. Colored curves are Ansatz solutions, Eq. 2.42. Gray vertical lines mark the edge level crossing points given by Eq. 2.48. (b) Auxiliary quantities. $\Delta$ is the discriminant in Eq. 2.45, only $\Delta > 0$ yields real solutions of edge energy $\varepsilon_{\pm}$ and is physical. $r_+:$ the ratio of Eq. 2.39 corresponding to the $\varepsilon_+$ branch. It is real if $\Delta > 0$. $r_-$ can be obtained from Eq. A.3 and is not plotted here. $\rho(\varepsilon_{\pm})$: the inter-edge tunnelling strength as defined in Eq. 2.24. These are solved in a self-consistent fashion in Appendix A. $\rho \to 0$ for open boundary, $\rho > 1$ for “enhanced-tunnelling” boundary. The transition $\rho = 1$ corresponds to bulk modes.
the edge solution merges into the bulk ($|\psi\rangle$ becomes extended instead of localized).

Fig. 2.5 shows the Ansatz solutions (colored curves) vs. the numerical open boundary spectrum (circular dots) in 2.5a, and auxiliary variables $\Delta$, $r(\varepsilon_+)$ and $\rho(\varepsilon_\pm)$ in 2.5b. We note that:

1. $\Delta < 0$ for $|k_1| \lesssim 0.125$. In this region the Ansatz does not yield real energy solutions.

2. For $|k_1| \gtrsim 0.125$, the Ansatz yields physical solutions. One then computes $\rho$ self-consistently; $\rho \ll 1$ means open boundary, whereas $\rho > 1$ means enhanced-tunneling boundary. The transition happens when the open boundary edge modes merge with the bulk bands. The $\varepsilon_+$ branch (red curve) has open boundary for $k_1 \in [-\pi, -0.62\pi] \cup [-0.17\pi, -0.14\pi] \cup [0.62\pi, \pi]$, while the $\varepsilon_-$ branch (blue curve) has open boundary for $k_1 \in [-\pi, -0.48\pi] \cup [0.49\pi, \pi]$. In these regions, the Ansatz solutions overlap with the open-boundary numerics (filled circles). Note that the $\varepsilon_+$ branch briefly becomes open boundary in $[-0.17\pi, -0.14\pi]$. Without the Ansatz solution, one would have taken it to be part of the bulk spectrum.

3. Within the physical regime ($\Delta > 0$), the two branches cross twice, marked by the two vertical gray lines, one with $\rho(\varepsilon_+) \to 0$ and the other with $\rho(\varepsilon_+) > 1$. In both cases $\rho(\varepsilon_+) = \rho(\varepsilon_-)$. These two edge crossing $k_1$ points are described by Eq. 2.48 to be discussed later. While only the one with $\rho \to 0$ is relevant for the open-boundary edge spectrum, we shall see in the next section that both have geometrical significance.

### 2.2.4 Edge modes crossing points

The hallmark of quantum Hall systems is the existence of counter-propagating edge states at opposite edges, which in the energy band structure will interpolate between bands separated by the bulk gap. The existence of the spectral flow implies that the two edge states must become degenerate at some $k_1$ point. With the

---

$^2$Eq. 2.42 implies that $r_\pm \in \mathbb{R}$ is sufficient for $\varepsilon_\pm \in \mathbb{R}$. Eq. A.5 implies that $\varepsilon_\pm \in \mathbb{R}$ is sufficient for $r_\pm \in \mathbb{R}$, hence the equivalence.
edge solution at hand, we can identify the exact location of these crossing points. The condition for $\varepsilon_+ = \varepsilon_-$ is that

$$p_2(h_1 - h_2) + 2p_1 \text{Re}(v_2 - v_1) = 0,$$

see Appendix A for derivation. For the zigzag edge, this implies

$$m + 2 \sin \phi \sin k_c(t_1 + t_2 + t_3) = 0 \quad (2.47)$$

whence

$$k_c = \begin{cases} -\sin^{-1} \frac{m}{2(t_1 + t_2 + t_3) \sin \phi} \\ \pi + \sin^{-1} \frac{m}{2(t_1 + t_2 + t_3) \sin \phi} \end{cases}$$

(2.48)

Here $k_c$ denotes the values of $k_1$ where the two edge modes are energetically degenerate. The bearded edge solution is obtained by switching the 1 and 2 subscripts in $p_i, h_i$ and $v_i$, for which Eq. 2.46 implies

$$\left(m + 2t_1 \sin \phi \sin k_c \right) \cos \frac{k_c}{2} + (t_2 + t_3) \sin \phi \sin \frac{k_c}{2} = 0. \quad (2.49)$$

This could be recast as a cubic equation for $\tan(k_c/2)$, but it does not afford a particularly simple closed form solution.

What is the significance of these edge crossing points? Recall that the bulk Hamiltonian $H(k_1, k_2)$ maps each $k$ to a $B(k)$ vector. Fixing $k_1$ while varying $k_2$ will drive the $B$ vector along a closed curve in 3D. It turns out that at both $k_c$, this curve lies on a plane passing through the origin. To see this, we note that at the edge crossing point, the $2 \times 2$ blocks $h(k_1)$ and $v(k_1)$ are related via

$$h = \varepsilon_c \mathbb{I} + \frac{p_1}{p_2} (v + v^\dagger)$$

(2.50)

where $\varepsilon_c = \varepsilon_+(k_c) = \varepsilon_-(k_c)$ (cf. Eq. A.8). The bulk Hamiltonian at the edge crossing points is then

$$H(k_c, k_2) = h + v e^{ik_2} + v^\dagger e^{-ik_2} = v \left( \frac{p_1}{p_2} + e^{ik_2} \right) + \text{H.c.} + \varepsilon_c \mathbb{I},$$

(2.51)
and the corresponding $B(k_c, k_2)$ is

$$
B_z = \left( \frac{p_1}{p_2} + \cos k_2 \right) \text{Re} (v_1 - v_2) - \sin k_2 \text{Im} (v_1 - v_2), \\
B_x = p_1 + p_2 \cos k_2, \quad B_y = p_2 \sin k_2,
$$

(2.52)

It is then easy to check that $B$ is a linear combination of two independent vectors,

$$
B(k_c, k_2) = B_x (\hat{x} - \text{Re}(q) \hat{z}) + B_y (\hat{y} - \text{Im}(q) \hat{z})
$$

(2.53)

where

$$
q = \frac{v_2 - v_1}{p_2}
$$

(2.54)

is independent of $k_2$. The path of $B(k_c, k_2)$ is thus coplanar and normal to the vector

$$
\mathbf{n} \equiv \text{Re}(q) \hat{x} + \text{Im}(q) \hat{y} + \hat{z}.
$$

(2.55)

It is easy to verify that the origin itself is on the plane.

An interesting observation is that in graphene, the bulk $2 \times 2$ Hamiltonian is always off-diagonal, $H(k) = \begin{pmatrix} 0 & \gamma_k \\ \gamma_k & 0 \end{pmatrix}$ where $\gamma_k = 1 + e^{ik_1} + e^{ik_2}$, thus it is coplanar at any $k_1$ value. The origin is inside the path of $B$ for $|k_1| > 2\pi/3$, and outside otherwise, hence the well known result that its two zigzag edge modes are degenerate at $\varepsilon = 0$ for $|k_1| > 2\pi/3$, as we verified in §2.1.2. For the bearded edge, the degenerate edge modes connect the two Dirac points in the other way, namely with $|k_1| < 2\pi/3$.

### 2.2.5 Relation to 1D topological insulators

In one spatial dimension, topological insulators are characterized by their Berry phase $\gamma$, where $\gamma = \pi$ signifies a nontrivial band topology and $\gamma = 0$ trivial. For example, the Su-Schrieffer-Heeger model [63] describes a 1D tight binding chain of alternating strong and weak hopping parameters $t_1$ and $t_2$, and in $k$-space its Hamiltonian is $H(k) = \begin{pmatrix} 0 & \gamma_k \\ \gamma_k & 0 \end{pmatrix}$ where $\gamma_k = t_1 + t_2 e^{ik}$. $H(k) = b_k \cdot \sigma$ is off-diagonal hence its polarization vector $b_k$ moves in the $xy$ plane. Its Berry phase is $\pi$ if $\gamma_k$
winds once around the origin, \( |t_1| < |t_2| \), for which case an open chain\(^3\) would host two zero modes, one at each end. Incidentally, the end states of the SSH model can be solved by the same method as we used for the Haldane model.

The Haldane model has two coplanar points on the \( k_x \) axis, and the Berry phases there must be either \( \pi \) or 0. In this case a \( \pi \) Berry phase implies the existence of *degenerate* edge states (not necessarily at zero energy), whereas 0 implies their absence. Since the Chern number is the winding number of the Berry phase over \( k_x \), a nonzero Chern number requires the Berry phase at the two coplanar points to be different. The topological classification of the 2D Haldane model thus reduces to the classification of two 1D problems at the coplanar points.

The reduction of classification of higher dimensional problems to lower dimensions very general. For example, for 2D time reversal invariant (TRI) \( \mathbb{Z}_2 \) topological insulators, one can focus on TRI \( k_x \) points at \( k_x = 0 \) or \( \pi \) and look for the non-Abelian Berry phases there. We shall have more to say about the non-Abelian Berry phase when studying Wannier functions in Chapter 4.

\subsection{2.2.6 Application to edge/end states of other models}

Before closing, we note that the method developed here for solving the zigzag edge states in the Haldane model can be applied to other two band models in 2D and 1D directly. In Appendix A, we show this for models of graphene and boron nitride. Other models include the Su-Schrieffer-Heeger model [63], Kitaev’s superconducting chain [39], and Creutz ladder model [18] in 1D.


---

\(^3\)Crucially, the open chain should be obtained by severing a periodic chain on a link which connects *different* unit cells. Severing a link within a unit cell is equivalent to exchanging \( t_1 \) and \( t_2 \) and reverses the topological classification.
Chapter 3

Quantum entanglement

The most general description of a quantum state is by means of a
density matrix $\rho$. For example, if an experiment prepares an ensemble $\{\ket{\Psi_n}\}$ with probabilities $\{p_n\}$ where $\sum_n p_n = 1$, then the density matrix is an incoherent superposition, $\rho \equiv \sum_n p_n \ket{\Psi_n} \bra{\Psi_n}$, and the expectation value of any physical quantity $O$ is $\langle O \rangle \equiv \sum_n p_n \bra{\Psi_n} O \ket{\Psi_n} = \text{Tr}(\rho O)$. Different $\ket{\Psi_n}$s are not required to be mutually orthogonal, a relief for experimentalists. If $\rho$ is a projection onto a single state $\ket{\Psi}$, viz. $\rho^2 = \rho = \ket{\Psi} \bra{\Psi}$, then the system is in a pure state. Otherwise it is in a mixed state. The von Neumann entropy $S \equiv -\langle \log \rho \rangle$ provides a measure of how “mixed” a state is; for a pure state, $S = 0$.

A density matrix may arise as the Boltzmann factor $\rho \propto e^{-H/T}$ in quantum statistical mechanics, where $H$ is the Hamiltonian of the system and $T$ is temperature. In this case, the “mixedness” is due to thermal fluctuation. However, it may also arrive via a seemingly rather different route: beginning with a pure state $\rho = \ket{\Psi} \bra{\Psi}$, one can divide its degrees of freedom into two sets $A$ and $B$, then the subsystem consisting of $A$ is described by a reduced density matrix (RDM) $\rho_A = \text{Tr}_B(\rho)$ by tracing out $B$, in the sense that the expectation value of any physical quantity $O_A$, local to $A$, is $\langle O_A \rangle = \text{Tr}(\rho_A O_A)$. The von Neumann entropy of $\rho_A$, viz. $S_A = -\langle \log \rho_A \rangle$, measures the “mixedness” of $A$ as a remnant of obliterating the $B$ degrees of freedom: $A$ and $B$ are said to be entangled if $S_A \neq 0$.

Let us pause here and appreciate the peculiarity of quantum entanglement: even if we have complete knowledge over an entire system ($S = 0$), we may not
be able to describe a part of it with the same determinacy \((S_A \neq 0)\). This is in
drastic contrast with classical physics, where the knowledge of an entire system,
as a point in phase space, entails complete knowledge of any constituent part.

### 3.1 General formulation

To take a closer look at the problem of bipartite entanglement, consider a
pure state \(|\Psi\rangle\) in the Hilbert space \(\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B\), where \(\mathcal{H}_{A,B}\) are the Hilbert
spaces of the constituent \(A\) and \(B\) subsystems. Let \(|\{A_a\}\rangle\) and \(|\{B_i\}\rangle\) be arbitrary
orthonormal bases of \(\mathcal{H}_A\) and \(\mathcal{H}_B\), respectively. In general \(|\Psi\rangle\) has the expansion

\[
|\Psi\rangle = \sum_{a,i} \Psi_{a,i} |A_a\rangle \otimes |B_i\rangle .
\]  

(3.1)

It is possible to choose a pair of “optimal” bases for which each basis state in \(A\) is
coupled to at most one basis state in \(B\), by the so-called Schmidt decomposition:
First invoke singular value decomposition (SVD) on the coefficient matrix \(\Psi_{a,i}\),

\[
\Psi_{a,i} = \sum_{\mu=1}^{R} U_{a,\mu} \sqrt{\lambda_{\mu}} V_{\mu,i}^\dagger
\]  

(3.2)

where \(U\) and \(V\) are unitary transformations and \(\sqrt{\lambda_{\mu}} > 0\) are the singular values.
One can now rewrite

\[
|\Psi\rangle = \sum_{\mu} \sqrt{\lambda_{\mu}} |a_\mu\rangle \otimes |b_\mu\rangle
\]  

(3.3)

with

\[
|a_\mu\rangle \equiv \sum_a U_{a,\mu} |A_a\rangle \quad , \quad |b_\mu\rangle \equiv V_{\mu,i}^\dagger |B_i\rangle .
\]  

(3.4)

The singular value index \(\mu\) can be thought of as labeling entanglement channels
through which \(A\) and \(B\) are connected. The number of channels \(R\) (i.e. the number
of nonzero singular values) is bounded by the lesser dimension of the two Hilbert
subspaces,

\[
R \leq R_{\text{max}} \equiv \text{Min}[\text{Dim}(\mathcal{H}_A), \text{Dim}(\mathcal{H}_B)] ,
\]  

(3.5)
thus the orthonormal sets \{\ket{a_{\mu}}\} and \{\ket{b_{\mu}}\} are not necessarily complete in their respective Hilbert spaces, which is okay because their orthogonal complements, 
\([\text{Span}\{\ket{a_{\mu}}\}]^\perp\) and \([\text{Span}\{\ket{b_{\mu}}\}]^\perp\), are not entangled. In terms of RDMs, one has 
\[
\rho_A = \sum_{\mu} \lambda_{\mu} \ket{a_{\mu}}\bra{a_{\mu}}, \quad \rho_B = \sum_{\mu} \lambda_{\mu} \ket{b_{\mu}}\bra{b_{\mu}},
\]  
which have identical nonzero eigenvalues. From this point of view, Schmidt decomposition of a pure state provides a partial map between the two complementary Hilbert spaces, taking \(\ket{a_{\mu}} \in \mathcal{H}_A\) to \(\ket{b_{\mu}} \in \mathcal{H}_B\).

The RDM eigenvalues \(\lambda_{\mu}\) form the *entanglement spectrum*, from which one obtains the entanglement entropy defined as the von Neumann entropy of \(\rho_A\) (equivalently of \(\rho_B\)),
\[
S_A = S_B = - \sum_{\mu=1}^{R} \lambda_{\mu} \log \lambda_{\mu}.
\]  
Clearly, \(S_A = S_B = 0\) if \(R = 1\), *i.e.* there is no entanglement in a product state \(\ket{\Psi} = \ket{a_1} \otimes \ket{b_1}\). For a fixed number of entanglement channels \(R\), maximal entanglement is obtained when \(\lambda_{\mu} = 1/R\) (with the normalization constraint \(\sum_{\mu} \lambda_{\mu} = 1\)). A global *maximal entanglement* is achieved when \(R\) is saturated, \(S_{\text{max}} = \log R_{\text{max}}\), cf. Eq. 3.5.

### 3.2 Real-space entanglement of a single particle wavefunction

A single particle wavefunction is the simplest quantum object one can manipulate. It may also describe some many-body physics, for example a Bose condensate. We analyze its entanglement with respect to a real space bipartition.

#### 3.2.1 Self entanglement of a single particle

Consider a particle of state \(\ket{\psi}\) on a 1D lattice of length \(2N\) and we want to know the entanglement between the left and right halves of the lattice. \(\ket{\psi}\) can
be written as a direct sum of the two halves,

$$|\psi\rangle = |\psi_L\rangle \oplus |\psi_R\rangle ,$$

(3.8)

notice that for a normalized $|\psi\rangle$, $|\psi_L\rangle$ and $|\psi_R\rangle$ are in general not normalized. In terms of column vectors,

$$\begin{pmatrix}
\psi_1 \\
\psi_2 \\
\vdots \\
\psi_N \\
\psi_{N+1} \\
\psi_{N+2} \\
\vdots \\
\psi_{2N}
\end{pmatrix} =
\begin{pmatrix}
\psi_1 \\
\psi_2 \\
\vdots \\
\psi_N \\
0 \\
0 \\
\vdots \\
0
\end{pmatrix} +
\begin{pmatrix}
0 \\
0 \\
\vdots \\
0 \\
\psi_{N+1} \\
\psi_{N+2} \\
\vdots \\
\psi_{2N}
\end{pmatrix}$$

(3.9)

In order to convert the direct sum into an ordinary sum, we need to introduce the vacuum for both halves, $|\emptyset_L\rangle$ and $|\emptyset_R\rangle$, representing a state with no particle in them. Eq. 3.8 can now be rewritten as

$$|\psi\rangle = |\psi_L\rangle \otimes |\emptyset_R\rangle + |\emptyset_L\rangle \otimes |\psi_R\rangle = \sqrt{\lambda_L}|\tilde{\psi}_L\rangle \otimes |\emptyset_R\rangle + \sqrt{\lambda_R}|\emptyset_L\rangle \otimes |\tilde{\psi}_R\rangle ,$$

(3.10)

which is already in a Schmidt-decomposed form. Here $|\tilde{\psi}_{L,R}\rangle$ are normalized versions of $|\psi_{L,R}\rangle$, which should be understood as $N$-dimensional column vectors, i.e., one-body states living in half system,

$$\lambda_i = \langle \psi_i | \psi_i \rangle , \quad |\tilde{\psi}_i\rangle = \frac{|\psi_i\rangle}{\sqrt{\lambda_i}} , \quad i = L, R ,$$

(3.11)

$$\lambda_L + \lambda_R = 1 .$$

(3.12)

It is curious that even the single particle entanglement problem should invoke the notion of a vacuum, an intrinsically many-body concept.

The many-body density matrix is

$$\rho = \lambda_L|\tilde{\psi}_L\rangle \langle \tilde{\psi}_L | \otimes |\emptyset_R\rangle \langle \emptyset_R | + \lambda_R|\emptyset_L\rangle \langle \emptyset_L | \otimes |\tilde{\psi}_R\rangle \langle \tilde{\psi}_R | + \sqrt{\lambda_L \lambda_R} \left(|\tilde{\psi}_L\rangle \langle \emptyset_L | \otimes |\tilde{\psi}_R\rangle \langle \emptyset_R | + h.c.\right) .$$

(3.13)
The reduced density matrix for the left half is then
\[ \rho_L = \text{Tr}_R(\rho) = \lambda_L \langle \tilde{\psi}_L | \tilde{\psi}_L \rangle + \lambda_R |\emptyset_L \rangle \langle \emptyset_L | , \tag{3.14} \]
and similarly for the right-half RDM. Physically, a left vacuum means the particle is entirely in the right half, so Eq. 3.14 simply says the particle has a probability of \( \lambda_L(\rho) \) to be in the left (right) half. Note that cross terms in Eq. 3.13 do not contribute to the RDM. The entanglement spectrum consists of the occupation numbers \( \lambda_L \) and \( \lambda_R = 1 - \lambda_L \). The entanglement entropy is
\[ S = -\langle \log \rho \rangle = -\lambda_L \log \lambda_L - \lambda_R \log \lambda_R , \tag{3.15} \]
which is formally identical to the classical entropy of mixing (thinking of \( L \) and \( R \) as labeling species of ideal gas particles).

**Second quantized form**

For later convenience, we now rewrite the above in the language of second quantization. The state of the full system is
\[ |\psi\rangle = b^\dagger |\emptyset\rangle , \quad \text{where} \quad b^\dagger \equiv \sqrt{\lambda_L} b_L^\dagger + \sqrt{\lambda_R} b_R^\dagger \,, \quad |\emptyset\rangle = |\emptyset_L\rangle \otimes |\emptyset_R\rangle . \tag{3.16} \]
Note that the last equation implies the absence of entanglement in the full vacuum.\(^1\)

The density matrix is thus
\[ \rho = b^\dagger |\emptyset\rangle \langle \emptyset | b = \lambda_L b_L^\dagger |\emptyset\rangle \langle \emptyset | b_L + \lambda_R b_R^\dagger |\emptyset\rangle \langle \emptyset | b_R + \sqrt{\lambda_L \lambda_R} \left( b_L^\dagger |\emptyset\rangle \langle \emptyset | b_R + b_R^\dagger |\emptyset\rangle \langle \emptyset | b_L \right) . \tag{3.17} \]

The reduced density matrix for the left half is
\[ \rho_L = \text{Tr}_R(\rho) = \langle \emptyset_R |\rho |\emptyset_R \rangle + \langle \emptyset_R | b_R^\dagger \rho b_R |\emptyset_R \rangle \\
= \lambda_L b_L^\dagger |\emptyset_L\rangle \langle \emptyset_L | b_L + \lambda_R |\emptyset_L\rangle \langle \emptyset_L | , \tag{3.18} \]
which is a sum of a 1-body projector and the vacuum projector on the left half.

Note that the partial trace automatically picks out *terms with balanced labels* –

\(^1\)More appropriately, we are *ignoring* any entanglement in the vacuum. Typically a quasi-particle vacuum is a many-body reference state which may be bipartite entangled in terms of the original particles.
terms with the same number of $L$ (and of $R$) on the two sides of $|\emptyset\rangle\langle\emptyset|$. The coefficients in front of the (mutually orthogonal) projectors constitute the entanglement spectrum, from which one can recover Eq. 3.15 as the entanglement entropy.

**Gaussian form**

It is possible to write $\rho_L$ in Eq. 3.18 above in a Gaussian form, provided $b_L \equiv c$ is a fermion mode. Assume the result is

$$\rho_L = \frac{\exp(\varepsilon c^\dagger c)}{\text{Tr} \exp(\varepsilon c^\dagger c)}.$$ (3.19)

In the occupation number basis one has

$$\rho_L = \frac{e^\varepsilon}{e^\varepsilon + 1} |1\rangle\langle 1| + \frac{1}{e^\varepsilon + 1} |0\rangle\langle 0|.$$ (3.20)

Comparing with Eq. 3.18, one can identify

$$\varepsilon = \log \frac{\lambda_L}{1 - \lambda_L}.$$ (3.21)

This is the single-fermion-mode limit of Eq. 3.48 below.

### 3.2.2 Entanglement of a Bose condensate

The second quantized formalism can easily be generalized to a Bose condensate. Consider $N$ bosons condensed onto the same single-particle state,

$$|\Psi\rangle = b^{iN} \frac{\sqrt{N!}}{\sqrt{N!}} |\emptyset\rangle.$$ (3.22)

What is the entanglement entropy between the left and right halves? One might guess that $S \equiv NS_1/\log(N!)$, where $S_1$ is the entanglement entropy of a single particle – the factor of $N$ would arise from “additivity” of entropy, and $N!$ would be the Gibbs factor accounting for the overcounting due to bosons being identical particles. As we shall see, the actual answer is obtained by replacing the Gibbs factor with $\left(\frac{\bar{n}}{\bar{n}}\right)$ where $\bar{n}$ is the average boson occupancy in the left half, that is, left and right are distinguishable.
The density matrix of the condensate is
\[ \rho = |\Psi\rangle\langle\Psi| = \sum_{n=0}^{N} \binom{N}{n} \lambda_L^n \lambda_R^{N-n} P_n^{(L)} \otimes P_{N-n}^{(R)} + \text{terms with unbalanced labels}, \tag{3.23} \]

where \( P_n^{(L)} \) is the projector in the left half onto the normalized state \( |n^{(L)}\rangle \) with \( n \) bosons in the state \( b_L^\dagger |\emptyset_L\rangle \) (implying \( N - n \) in the right half with state \( b_R^\dagger |\emptyset_R\rangle \)),
\[ P_n^{(i)} = |n^{(i)}\rangle\langle n^{(i)}|, \quad |n^{(i)}\rangle = \frac{b_i^{\dagger n} |\emptyset_i\rangle}{\sqrt{n!}}, \quad i = L, R \tag{3.24} \]

Terms with unbalanced labels are like those in the bracket of Eq. 3.17, which will drop upon partial trace and thus will not contribute to entanglement spectrum/entropy. The RDM of the left half is
\[ \rho_L = \text{Tr}_R(\rho) = \sum_{n=0}^{N} \langle n^{(R)} | \rho | n^{(R)} \rangle = \sum_{n=0}^{N} \lambda_n P_n^{(L)} , \quad \lambda_n = \binom{N}{n} \lambda_L^n \lambda_R^{N-n} . \tag{3.25} \]

The entanglement spectrum \( \lambda_n \) has a binomial distribution, hence the corresponding entanglement entropy is known as a \textit{binomial entropy}. Explicitly,
\[ S = -\sum_{n=0}^{N} \lambda_n \log(\lambda_n) = -\sum_{n=0}^{N} \lambda_n \log \left( \frac{N}{n} \right) - N \left( \lambda_L \log \lambda_L + \lambda_R \log \lambda_R \right) . \tag{3.26} \]

The second term in \( S \) is the one-body part which is simply \( N \) times Eq. 3.15. The first term is always negative, implying it being an overcounting effect.

The binomial entropy Eq. 3.26 has a well known asymptotic form at large \( N \),
\[ S = \frac{1}{2} \log[2\pi e N \lambda_L \lambda_R] + O(N^{-1}) , \tag{3.27} \]
which follows from the so-called de Moivre-Laplace theorem. It scales as \( \log N \) instead of \( N \) which one might have expected from additivity. This can be understood in the following way: At large \( N \), the binomial distribution \( \lambda_n \) has a sharp peak near
\[ \bar{n} = N\lambda_L , \tag{3.28} \]
that is, the left half is dominated by the $\bar{n}$-particle sector. Eq. 3.28 can be derived by demanding $\lambda_{n+1}/\lambda_n = 1$ at $n = \bar{n}$. Assuming the effective width $w$ of $\lambda_n \log \lambda_n$ is the same as $\lambda_n$, \textit{viz.}

$$w = \frac{1}{\lambda_{\bar{n}}}$$

(3.29)

then the total entropy can be estimated as

$$S \simeq -w \lambda_{\bar{n}} \log \lambda_{\bar{n}} = -\log \lambda_{\bar{n}} = -\log \left( \frac{N}{\bar{n}} \right) - \left[ \bar{n} \log \frac{\bar{n}}{N} + (N - \bar{n}) \log \frac{N - \bar{n}}{N} \right],$$

(3.30)

where $\binom{N}{\bar{n}}$ should be computed using Gamma functions since $\bar{n}$ is not an integer,

$$\binom{N}{\bar{n}} = \frac{\Gamma(N+1)}{\Gamma(\bar{n}+1)\Gamma(N-\bar{n}+1)}.$$  

(3.31)

Thinking of $e^S$ as the number of microscopic ways through which to realize a macroscopic state, and noting that

$$e^S = \left[ e^{S_1} \right]^{N \binom{N}{\bar{n}}},$$

(3.32)

the deviation from additivity comes from the overcounting $\binom{N}{\bar{n}}$ of $\bar{n}$ bosons in the left half due to them being identical particles.

A naive application of Stirling’s formula to evaluate Eq. 3.30 would simply yield zero. Instead, we use the asymptotic expression

$$\Gamma(z) = z^{z-\frac{1}{2}} e^{-z} \sqrt{2\pi} \left[ 1 + \frac{1}{12z} + O(z^{-2}) \right].$$

(3.33)

Then

$$\log \Gamma(z+1) = \left( z + \frac{1}{2} \right) \log z - z + \frac{1}{2} \log(2\pi) + \frac{7}{12z} + O(z^{-2})$$

(3.34)

from which

$$S \simeq \frac{1}{2} \log \frac{\bar{n}(N - \bar{n})}{N} + \frac{1}{2} \log(2\pi) + O(N^{-1}) = \frac{1}{2} \log[2\pi N \lambda_L \lambda_R].$$

(3.35)

This crude estimation (originating from Eq. 3.29) is only off by a constant 0.5 below the numerical value for all $N$, which is also the difference from Eq. 3.27.
3.3 Real-space entanglement of free fermions

The partial trace involved in the construction of a reduced density matrix can be turned into a path integral. However, if the trace is over fermionic degrees of freedom, the resulting path integral is Grassmannian and is hard to evaluate numerically. For free fermions, Peschel [49] found a way to derive the RDM without evaluating the Grassmannian integral. We review Peschel’s method in this section.

3.3.1 Relation between a Gaussian density matrix and its one-body correlation matrix

A Gaussian density matrix $\rho$, where $\log \rho$ is bilinear in the creation and annihilation operators, is completely determined by its correlation matrix $G$, which itself is essentially a single-particle “density matrix”, in the sense that for any single-particle quantity $O$, its expectation value is $\text{Tr}(G O)$. Below we derive the connection between $\rho$ and $G$.

Consider a many-body density matrix $\rho$ (un-normalized),

$$\rho = e^{-\hat{\Gamma}} .$$

(3.36)

For example, if $\hat{\Gamma} = \beta \hat{H}$, then $\rho$ is the thermal density matrix. Assume $\hat{\Gamma}$ has the most general fermionic bilinear form,

$$\hat{\Gamma} = A_{ij} \psi_i^\dagger \psi_j + \frac{1}{2} (B_{ij} \psi_i^\dagger \psi_j^\dagger + \text{h.c.}) = \frac{1}{2} \Psi^\dagger \Gamma \Psi + \frac{1}{2} \text{Tr} A .$$

(3.37)

with

$$\Psi \equiv (\psi_1, \psi_2, \ldots, \psi_N, \psi_1^\dagger, \psi_2^\dagger, \ldots, \psi_N^\dagger)^T , \quad \{\Psi_i, \Psi_j^\dagger\} = \delta_{i,j}$$

(3.38)

$$\Gamma \equiv \begin{pmatrix} A & B \\ B^\dagger & -A^T \end{pmatrix} = \begin{pmatrix} A & B \\ -B^* & -A^* \end{pmatrix} ,$$

(3.39)

where $\psi_i$ are fermions and $N$ is the number of lattice sites. The last equality follows from $A^\dagger = A$ (hermiticity of $\hat{\Gamma}$) and $B^T = -B$ (fermion anticommutation). By construction, the matrix $\Gamma$ has a particle-hole symmetry: if $(v^u)$ is an eigenvector
of $\Gamma$ with eigenvalue $\varepsilon$, then $(v^* u)$ is also an eigenvector with eigenvalue $-\varepsilon$. Thus a unitary $S$ which diagonalizes it has the form

$$S = \begin{pmatrix} U & V^* \\ V & U^* \end{pmatrix}$$

(3.40)

such that

$$S^\dagger \Gamma S = \begin{pmatrix} E \\ -E \end{pmatrix}, \quad E = \text{diag}(\varepsilon_1, \varepsilon_2, \cdots \varepsilon_N).$$

(3.41)

The diagonal basis is

$$\Phi = S^{-1} \Psi = (\phi_1, \phi_2, \cdots, \phi_N, \phi_1^\dagger, \phi_2^\dagger, \cdots, \phi_N^\dagger)^T, \quad \{\phi_a, \phi_b^\dagger\} = \delta_{a,b}$$

(3.42)

in terms of which $\hat{\Gamma}$ becomes

$$\hat{\Gamma} = \frac{1}{2} \sum_{a=1}^N \varepsilon_a (\phi_a^\dagger \phi_a - \phi_a \phi_a^\dagger) + \frac{1}{2} \text{Tr} A = \sum_{a=1}^N \varepsilon_a \phi_a^\dagger \phi_a + \frac{1}{2} \left( \text{Tr} A - \sum_{a=1}^N \varepsilon_a \right).$$

(3.43)

The constant term vanishes if the matrix $B = 0$.

The density matrix in the diagonal basis can be factorized,

$$\rho = \prod_a \tilde{\rho}_a / Z_a, \quad \tilde{\rho}_a = \exp(-\varepsilon_a \phi_a^\dagger \phi_a), \quad Z_a = \text{Tr} \tilde{\rho}_a = 1 + e^{-\varepsilon_a},$$

(3.44)

where the partition functions $Z_a$ are easily evaluated in the occupation number basis of $\phi_a^\dagger \phi_a$. We have restored normalization for $\rho$. With factorization, one can easily verify that

$$\langle \phi_a \phi_b^\dagger \rangle = \frac{\delta_{a,b}}{e^{-\varepsilon_a} + 1}, \quad \langle \phi_a \phi_b \rangle = \langle \phi_a^\dagger \phi_b^\dagger \rangle = 0.$$ (3.45)

The correlation matrix in the $\Phi$ basis is defined as (note that $\Phi \Phi^\dagger$ is a matrix of operators)

$$G^\Phi \equiv \langle \Phi \Phi^\dagger \rangle = \begin{pmatrix} 1 \exp(-E) + 1 \\ \exp(E) + 1 \end{pmatrix} = \frac{1}{\exp(-S^\dagger \Gamma S) + 1},$$

(3.46)

which is diagonal. The last term follows from Eq. 3.41. Transforming back to the $\Psi$ basis yields

$$G \equiv \langle \Psi \Psi^\dagger \rangle = S G^\Phi S^\dagger = \frac{1}{\exp(-\Gamma) + 1} = 1 - \frac{1}{\exp \Gamma + 1}$$

(3.47)
and its inverse relation
\[ \Gamma = \log \frac{G}{1 - G}. \] (3.48)

This is a generalization of the single mode result Eq. 3.21.

If \( \hat{\Gamma} \) conserves particle number, then the anomalous term \( B \) in Eq. 3.37 is absent. In such cases, denoting\[
G_{ab}^{\phi} \equiv \langle \phi_a \phi_b^\dagger \rangle, \quad G_{ij} \equiv \langle \psi_i \psi_j^\dagger \rangle,
\] (3.49)
one has\[
1 - G = \frac{1}{\exp(A) + 1} \iff A = \log \frac{G}{1 - G}. \] (3.50)

Note that at zero temperature, \( 1 - G \) is a projection onto (single-particle) energy eigenstates below Fermi energy,
\[
1 - G = \sum_{\varepsilon_a < E_F} |\phi_a\rangle \langle \phi_a|.
\] (3.51)

### Entropy

Denote \( \rho_a = \tilde{\rho}_a / Z_a \) as the normalized density matrix for the diagonal modes. Then the total entropy is the sum of entropies of all modes,
\[
S = - \text{Tr} \left[ \left( \prod_a \rho_a \right) \sum_a (\log \rho_a) \right] = \sum_a S_a,
\] (3.52)

where
\[
S_a = - \text{Tr} (\rho_a \log \rho_a) = - f_a \log f_a - (1 - f_a) \log (1 - f_a)
\] (3.53)
and \( f_a \) are the Fermi weights,
\[
f_a \equiv f(\varepsilon_a) = \frac{1}{e^{\varepsilon_a} + 1}. \] (3.54)

Note the similarity between Eqs. 3.53 and 3.15.
3.3.2 Single-particle entanglement spectrum

Eq. 3.48 relates the coefficient matrix of log $\rho$ with the correlation matrix $G$. Peschel noted that the correlation matrix of a subsystem $G_{\text{sub}}$, computed using the corresponding reduced density matrix, must be identical to the same calculation using the full density matrix, thus $G_{\text{sub}}$ is simply a submatrix of the full correlation $G$. To compute entanglement spectrum of a free fermion system, one first obtains its full correlation matrix from its full density matrix, then take its submatrix, in an appropriate basis, according to the entanglement bipartition. In principle, one can then use Eq. 3.48 to obtain the reduced density matrix whose eigenvalues constitute the (many-body) entanglement spectrum. In practice, it is sufficient to compute the spectrum of $G_{\text{sub}}$, which contains the same information as the many-body entanglement spectrum. The spectrum of $G_{\text{sub}}$ is known as the single-particle entanglement spectrum, and will be used later to study topological band systems.
Chapter 4

Wannier states

One of the defining features of quantum mechanics is Heisenberg’s Uncertainty Relation. As a consequence, localizations in momentum and coordinate spaces are mutually exclusive. One can see this from the projector onto a momentum eigenstate, $P_k \equiv |k\rangle\langle k| = \frac{1}{N} \sum_{x,x'} e^{ik(x-x')} |x\rangle\langle x'|$, where $|k\rangle = \frac{1}{\sqrt{N}} \sum_x e^{ikx} |x\rangle$. $P_k$ is diagonal in the $k$ representation, but fully spreaded in the $x$ representation. The identity operator, on the other hand, escapes this fate by definition: it is diagonal in both bases, $\mathbb{I} = \sum_k |k\rangle\langle k| = \sum_x |x\rangle\langle x|$. This is because the off-diagonal elements in the $x$ basis oscillate rapidly as $k$ changes and cancel each other over summation. One could imagine dividing all degrees of freedom in one basis (say $k$) into several groups, and ask how does localization in the conjugate basis emerge as different groups of DOFs are summed in sequence. This question is vague because the answer of course would depend on how one divides $k$. However, some natural choice arises when one considers a lattice system, wherein due to the breaking of continuous translational symmetry, the full space decomposes into one space labeled by crystal momentum $k$ and the other by band index $n$ (Bloch’s Theorem), thus one may choose to group the DOFs by the band index. Indeed, as we shall see in Chapter 5, band projectors themselves, as well as their sums, are fairly localized in coordinate space. In order to study phenomena associated with their localization – for example their adiabatic pumping driven by some external parameter – it is desirable to have a set of basis states within the projected subspace, which are localized, analogous to the $\{|x\rangle\}$ basis in the full space. Such localized eigenstates
of band projectors are known as Wannier states.

### 4.1 Bloch Theorem on a lattice

Consider a 1D periodic system consisting of \( N \) unit cells each with \( Q \) internal degrees of freedom (spin, orbital, sublattice, etc.). According to Bloch’s theorem, translational invariance on the scale of unit cells implies that its Hamiltonian can be block-diagonalized by Fourier transform,

\[
H = \sum_k P_k \otimes H_k \quad , \quad P_k = |k\rangle\langle k| \quad , \quad |k\rangle \equiv \frac{1}{\sqrt{N}} \sum_x e^{ikx} |x\rangle \quad (4.1)
\]

The \( N \times N P_k \) acts on the coordinate space (position of unit cells), and the \( Q \times Q \) \( H_k \) is the Fourier transform of \( H \) acting on internal space. The eigenstates of \( H \) are

\[
|\Psi_k^n\rangle = |k\rangle \otimes |\psi_k^n\rangle \quad , \quad H_k|\psi_k^n\rangle = E_k^n |\psi_k^n\rangle \quad , \quad n = 1, 2, \cdots Q . \quad (4.2)
\]

### 4.2 Wannier states of a single band

To get some intuition, consider first a single occupied band of an infinite chain. We will solve the eigenvalue problem of \( G\hat{\mathcal{X}}G \) \([40, 52]\). Here, \( G \equiv \sum_k |\Psi_k\rangle\langle \Psi_k| \) is the projector onto the filled band, cf. Eq.4.2 (we have dropped the band index). \( \hat{\mathcal{X}} = \sum_x x|x\rangle\langle x| \) is the position operator. The eigenstate must be a linear combination of states in the filled band,

\[
|\lambda\rangle = \sum_k f_k |k\rangle \otimes |\psi_k\rangle \quad , \quad G\hat{\mathcal{X}}G|\lambda\rangle = \lambda|\lambda\rangle \quad , \quad (4.3)
\]

In component form the eigenvalue equation reads

\[
\sum_{k'} \langle k|\hat{\mathcal{X}}|k'\rangle \times f_{k'} \langle \psi_k|\psi_{k'}\rangle = \lambda f_k \quad , \quad (4.4)
\]
note that $\hat{X}$ acts only on $|k\rangle$ but not on $|\psi_k\rangle$. It is easy to verify in the coordinate basis that

$$\hat{X}|k\rangle = -i\partial_k|k\rangle .$$

(4.5)

Thus integrating by parts once, we have

$$i\partial_k f_k = (\lambda - A_k) f_k , \quad A_k \equiv \langle \psi_k | i\partial_k | \psi_k \rangle .$$

(4.6)

The Berry connection $A_k$ of the internal space arises naturally from compensating the momentum shift, due to $\hat{X}$, in the coordinate space. The solution is

$$f_k = e^{-ik\lambda} e^{i\theta(k)} , \quad \theta(k) \equiv \int_0^k A_\kappa d\kappa .$$

(4.7)

where we have chosen the normalization $f_0 = 1$. Periodicity in $k$ requires that $f_{2\pi} = f_0$, implying

$$\gamma \equiv \int_0^{2\pi} A_k d\kappa = 2\pi(\lambda - I) , \quad I \in \mathbb{Z} ,$$

(4.8)

where $\gamma$ is the Berry phase. This quantizes the eigenvalues $\lambda$,

$$\lambda_I = \langle \lambda_I | \hat{X} | \lambda_I \rangle = \frac{\gamma}{2\pi} + I .$$

(4.9)

Thus the Wannier states $\{|\lambda_I\rangle\}$ are equally spaced; the displacement of $|\lambda_I\rangle$ from the $I^{th}$ unit cell is given by the Berry phase $\gamma/2\pi$.

The Wannier states solved above can be generated by translating a reference state,

$$|\lambda_I\rangle = \hat{T}_{\lambda_I} |w\rangle$$

(4.10)

where $\hat{T}$ is a coordinate space translation,

$$\hat{T}_\delta |x\rangle = |x + \delta\rangle = \sum_k e^{-i(x+\delta)} |k\rangle ,$$

(4.11)

and the reference state is

$$|w\rangle = \sum_k |k\rangle \otimes |\tilde{\psi}_k\rangle , \quad |\tilde{\psi}_k\rangle \equiv e^{i\theta(k)} |\psi_k\rangle .$$

(4.12)
Traditionally, Wannier states are defined by replacing $|\tilde{\psi}_k\rangle$ with $|\psi_k\rangle$, which amounts to a different gauge choice. This will not change the location of the Wannier centers $\lambda_I$ ($\gamma$ being gauge invariant), but will change the wavefunction spread. The specialty of $|\tilde{\psi}\rangle$ is that it traces out a parallel path, 

$$\langle\tilde{\psi}_k|\partial_k|\tilde{\psi}_k\rangle = 0 . \quad (4.13)$$

### 4.3 Wannier states of multiple bands

We now return to a periodic chain with $N$ unit cells and $Q$ bands, and solve the eigenvalue problem of $G\hat{T}G$. Here $\hat{T}$ is the translation operator in $k$ space, and $G$ projects onto filled bands, 

$$\hat{T} = \exp(-i\delta \hat{X}) , \quad \delta = \frac{2\pi}{N} , \quad (4.14)$$

$$G = \sum_k P_k \otimes G_k , \quad G_k = \sum_{n=1}^\nu |\psi_n^k\rangle\langle\psi_n^k| , \quad (4.15)$$

thus

$$G\hat{T}G = \sum_{k,k'} (P_k\hat{T}P_{k'}) \otimes (G_k G_{k'}) = \sum_\ell |k_\ell\rangle\langle k_{\ell+1}| \otimes G_\ell G_{\ell+1} , \quad k_\ell \equiv \ell\delta . \quad (4.16)$$

In momentum space it has the following block form

$$G\hat{T}G = \begin{pmatrix}
\begin{array}{ccc}
0 & G_1 G_2 & \cdot \\
0 & G_2 G_3 & \cdot \\
& & \ddots \\
G_N G_1 & & 0
\end{array}
\end{pmatrix} \quad (4.17)$$

each block being a $Q \times Q$ matrix. Assume its eigenstate is $|\lambda\rangle = (|f_1\rangle, |f_2\rangle, \cdots, |f_N\rangle)^t$ where $|f_\ell\rangle$ is a $Q$-dimensional vector, then the eigenvalue equation $G\hat{T}G|\lambda\rangle = \lambda|\lambda\rangle$ becomes

$$|f_\ell\rangle = \lambda^{-1} G_\ell G_{\ell+1} |f_{\ell+1}\rangle , \quad |f_{N+1}\rangle \equiv |f_1\rangle . \quad (4.18)$$
Iteration then yields
\[ W_N |f_N⟩ = λ^N |f_N⟩ \] (4.19)

where
\[ W_N ≡ G_N G_1 G_2 \cdots G_N . \] (4.20)
is the \( Q \times Q \) Wilson loop matrix.

One can similarly define \( W_ℓ \) for loops initiating at different \( k_ℓ \). These will share the same eigenvalues with \( W_N \) following the spectral invariance of cyclic permutations: For two arbitrary square matrices \( A \) and \( B \) of the same (finite) dimension, the spectrum of \( AB \) is identical to that of \( BA \), including the degeneracy structure. Note that \( A \) and \( B \) do not have to be invertible. Using this theorem, the leading projector in \( W \) can be moved to the tail without changing its spectrum. To prove this, it is sufficient to show that their nonzero spectra are identical.\(^1\) Assume \( AB |ψ⟩ = λ |ψ⟩ \) and \( λ ≠ 0 \), then \( BA B |ψ⟩ = λB |ψ⟩ \), thus \( λ ∈ spec(AB) \implies λ ∈ spec(BA) \). To show the preservation of degeneracy, assume the contrary: \( |ψ_1⟩ \) and \( |ψ_2⟩ \) are degenerate orthonormal eigenstates of \( AB \), but \( B |ψ_1⟩ \) and \( B |ψ_2⟩ = ℓ^{-1} B |ψ_1⟩ \) are linearly dependent. Then
\[ B(|ψ_1⟩ - ℓ|ψ_2⟩) = 0 \implies |ψ_1⟩ = ℓ|ψ_2⟩ + |0⟩ \] where \( |0⟩ \in Kernel(B) \). Thus \( AB |ψ_1⟩ = ℓ AB |ψ_2⟩ = \ell |ψ_1⟩ = ℓλ |ψ_2⟩ + λ |0⟩ \). \( AB |ψ_2⟩ = λ |ψ_2⟩ \) then implies \( λ |0⟩ = 0 \). Since \( λ ≠ 0 \), one must have \( |0⟩ = 0 \), contradicting the linear independence of \( |ψ_1⟩ \) and \( |ψ_2⟩ \). The above argument shows that the nonzero spectrum of \( AB \) is a subset of that of \( BA \) with the correct degeneracy. The same applies in the reverse direction, thus their spectra are identical.

For \( ν \) filled bands, the \( Q \times Q \) matrix \( W_N \) has only \( ν \) nonzero eigenvalues, a consequence of the \( G_N \) projections on both ends. To obtain its non-zero part, denote
\[ ξ_k ≡ (|ψ_1^k⟩, |ψ_2^k⟩, \cdots, |ψ_ν^k⟩) \] (4.21)

\(^1\)For finite dimension, the number of zeros complements the rank. This is not true for infinite dimension. E.g., for fermions, \( spec(\overline{c}c) = spec(\overline{cc}^\dagger) = \{0, 1\} \), but for bosons, \( spec(a^\dagger a) = \{0, 1, 2, \cdots\} \) whereas \( spec(aa^\dagger) = \{1, 2, \cdots\} \). The nonzero spectra in both cases are identical.
then the \( \nu \) nonzero eigenvalues of \( W_N \) constitute the full spectrum of the \( \nu \times \nu \) matrix

\[
W_N \equiv \xi_N^\dagger W_N \xi_N = \xi_N^\dagger \xi_1 \cdot \xi_2^\dagger \cdots \xi_{N-1}^\dagger \xi_N ,
\]

(4.22)

where we have used \( G_k = \xi_k \xi_k^\dagger \). As \( N \to \infty \), each piece in \( W_N \) can be written as

\[
\xi_k^\dagger \xi_{k+\delta} = \exp \left[ -i \delta A(k) - O(\delta^2) \right] ,
\]

(4.23)

where \( A(k) \) is a \( \nu \times \nu \) Berry connection matrix for the filled subspace,

\[
A_{mn}(k) = \langle \psi_k^m | i \partial_k | \psi_k^n \rangle , \quad m, n = 1, 2, \ldots, \nu .
\]

(4.24)

Let us denote the eigenvalues of \( W_N \) as

\[
\lambda_a^N = \rho_a e^{-i \gamma_a} .
\]

(4.25)

In the continuum limit, \( W_N \) becomes unitary and \( \rho_a = 1 \),

\[
W_N = \exp \left[ -i \int_0^N A(k) dk \right] ,
\]

(4.26)

where the arrow points to the direction of increasing \( k \) in the path ordering. Comparing the form of \( \lambda_a \) with that of \( \hat{T} \), we identify

\[
\text{Wannier center drifts} = - \frac{1}{2\pi} \text{Arg} \left[ \text{Spec}(W_N) \right] ,
\]

(4.27)

which describes the deviation of Wannier centers from the corresponding unit cell boundary. For \( \nu = 1 \) it reduces to the single band case. For finite \( N \), \( \rho_a < 1 \) in general. Its deviation from 1 characterizes the coordinate-space spread of the corresponding wavefunction.

### 4.4 Relation with entanglement spectrum

As discussed in §3.3.2, the single-particle entanglement spectrum is the spectrum of \( G_{\text{sub}} \), which is the submatrix of \( G \) with its spatial indices restricted to a subsystem, say, \( x < M \). At zero temperature, \( G \) is the projector onto filled bands.
Formally one may write \( (G_{sub}^{0} 0) = RGR \), where \( R \) is a coordinate-space projector, \( R_{xx'} = \delta_{xx'} \Theta(x - M) \) where \( \Theta \) is the Heaviside step function. \( G_{sub} \) and \( RGR \) share the same non-zero eigenvalues. Consider the case of zero temperature, where \( G \) is a projector onto filled bands. From the spectral invariance of cyclic permutations (see proof above Eq. 4.21), \( RGR \) and \( GRG \) have the same spectrum. On the other hand, the spectrum of \( GXG \) are the Wannier centers. Since \( R = \Theta(X - M) \) can be viewed as a coarse-graining of \( X \) where any \( x < M \) counts as 0 and \( x \geq M \) counts as 1, the entanglement spectrum is a coarse-graining of the Wannier centers, where only the Wannier state localized near the \( M^{th} \) unit cell will contribute to an entanglement occupancy close to \( \frac{1}{2} \). Wannier states localized at \( x < M \) will contribute \( f \sim 0 \), while those localized at \( x > M \) will contribute \( f \sim 1 \).

Chapter 5

Entanglement spectrum and
Wannier center flow in the
Hofstadter problem

In this chapter we revisit the celebrated Hofstadter model, which is a paradigmatic model for integer quantum Hall effect on a square lattice. Hatsugai pointed out [27] that on a cylindrical geometry, the edge states of the Hofstadter model exhibit spectral flow as a function of the momentum in the circular direction. This is an example bulk boundary correspondence. We will examine the entanglement spectrum and Wannier center flow in the Hofstadter model, and show that both exhibit similar spectral flows.

5.1 The Hofstadter problem

Hofstadter was the first to study the problem of hopping electrons on a square lattice in the presence of a uniform magnetic field, a lattice version of the Landau quantization. Plotting the energy spectrum as a function of the magnetic flux per elementary plaquette $\Phi$ led him to the discovery of the famous “Hofstadter’s butterfly”.

The tight-binding Hamiltonian of the Hofstadter problem is
\[
\hat{H} = - \sum_{\langle ij \rangle} t_{ij} c_i^\dagger c_j + h.c. \tag{5.1}
\]
where \( t_{ij} \) is a complex hopping amplitude from site \( i \) to a neighboring site \( j \),
\[
t_{ij} = |t_{ij}| \exp(iA_{ij}). \tag{5.2}
\]
We shall only consider \(|t_{ij}| = 1\). The hopping phase \( A_{ij} \) can be viewed as the vector potential integrated over a lattice constant. According to the path integral formalism, the phase accumulation of the electron around an elementary plaquette is \( \phi = S/h \) where \( S = \frac{e}{c} \int \mathbf{v} \cdot dA = \frac{e}{c} \oint A \cdot dl = \frac{e\Phi}{c} \) is the classical action arising from the coupling of the electron to the vector field \( \mathbf{A} \). In terms of \( A_{ij} \), one has
\[
\phi = \sum_{\langle ij \rangle \in \partial \square} A_{ij} = \frac{e\Phi}{hc} = 2\pi \Phi/\Phi_0, \quad \Phi_0 \equiv \frac{hc}{e} \tag{5.3}
\]
where \( \Phi_0 \) is the Dirac flux quantum, and \( \partial \square \) denotes the (directed) boundary links of an elementary plaquette.

Although the magnetic field itself is translationally invariant, the gauge field \( A_{ij} \) is not, hence in general \( \hat{H} \) cannot be block diagonalized by translational symmetry. However, when \( \Phi/\Phi_0 \) is a rational number – viz., \( \phi = 2\pi p/q \) for co-prime integers \( p \) and \( q \) – the phase factor arising from going around \( q \) plaquettes becomes \( e^{iq\phi} = 1 \), and so it is possible to arrange for the configuration of \( A_{ij} \) to be periodic over a magnetic unit cell (MUC) consisting of \( q \) elementary plaquettes. The bulk energy spectrum will accordingly split into \( q \) subbands. Two remarks are in order: (i) For irrational \( \phi/2\pi \), it is possible to approximate it with a rational number \( p/q \) to arbitrary precision; (ii) Consider the difference between \( \frac{p}{q} = \frac{1}{3} \) and \( \frac{100}{301} \). Physically they are almost identical, but mathematically one has 3 subbands, the other has 301. What happens is that as \( \phi/2\pi \) changes from \( \frac{1}{3} \) to \( \frac{100}{301} \), each of the three subbands roughly speaking splits further into \( \sim 100 \) subbands by opening up mini gaps. The proliferation of such mini gaps as \( \phi/2\pi \) progresses towards irrational numbers lies at the heart of the fractal structure in the Hofstadter butterfly.

Different shapes of MUCs are related via a gauge transformation. We choose the MUC to be a tower of \( 1 \times q \) plaquettes (i.e. along the \( y \) direction), and use the
Landau gauge

\[ A_{ij} = \phi y_i \delta_{x_i, x_j+1} \delta_{y_i, y_j}, \]  

(5.4)

where \((x_i, y_i)\) denote the \(x\) and \(y\) coordinates of site \(i\). Such an arrangement, *modulo* \(2\pi\), is periodic on the scale of the MUC. We shall always assume the \(x\) direction to be periodic. Then after Fourier transform in the \(x\) direction, the Hamiltonian matrix becomes

\[
H(k_x, N_y, z) = \begin{pmatrix}
2 \cos(k_x + \phi) & 1 & 0 & \cdots & z^* \\
1 & 2 \cos(k_x + 2\phi) & 1 & 0 \\
0 & 1 & \ddots & \ddots \\
\vdots & & & 1 \\
z & 0 & \cdots & 1 & 2 \cos(k_x + N_y \phi)
\end{pmatrix}
\]  

(5.5)

where \(N_y\) is the number of sites in the \(y\) direction, and the complex number \(z\) controls the boundary condition along \(y\): (i) if \(z = 0\), one has open boundaries in \(y\), *i.e.* the full system is put on a cylindrical geometry (\(x\) being periodic); (ii) if \(z = 1\), and \(N_y = K_y q\) for integer \(K_y\), there are \(K_y\) MUCs, and one has periodic boundary conditions in \(y\); the system has a toroidal geometry; (iii) unimodular complex \(z\) can be used to implement flux threading through the compactified \(y\) direction. Note that for both (i) and (ii), \(H\) is real and hence its eigenstates can be made real.

Top panels of Fig. 5.1 plot the energy spectrum of the Hofstadter model on a cylindrical geometry, with flux per plaquette \(p/q = 3/7\), using two different \(N_y\)s.

### 5.2 Entanglement spectrum

Imagine partitioning the Hofstadter lattice into two subsystems \(L\) (for lower) and \(U\) (for upper), where \(L\) is restricted to \(1 \leq y \leq M\) and \(U\) to \(M + 1 \leq y \leq N_y\). We shall refer to the interface of such bipartites as the *entanglement cut*. In this case, the cut is parallel to the \(x\) direction and preserves its translational symmetry, thus \(k_x\) is still a good quantum number. The entire entanglement spectrum can thus be organized by \(k_x\) much like the energy spectrum.
Figure 5.1: Energy and entanglement spectrum of the Hofstadter model with $p/q = 3/7$ and three filled bands. Top panels plot the energy levels, bottom panels plot the entanglement occupancies. These are obtained numerically for the square lattice Hofstadter model as a function of the conserved crystal momentum $k_x$. The lattice is put on a cylindrical geometry. The Fermi level $\mu$ lies inside the third gap (grey line), and the occupied bands below $\mu$ contribute a total Chern number $C = 1$. Top panels plot energy spectrum, with the black dots indicating bulk levels, the red and blue lines indicating edge levels localized along the lower $(y = 1)$ and upper $(y = N_y)$ edges, respectively. The vertical gray line marks the value of $k_x$ where the lower edge mode crosses the Fermi level. Lower panels plot the entanglement occupancies $f_a$ computed for the lower half of the cylinder $(1 \leq y \leq M)$, color and symbol-coded according to $a$. Although the overall flow appears continuous, there is a discontinuity in the occupancies $f_a(k_x)$ where the lower edge mode crosses $\mu$, resulting in a sudden color change in the plot.
Figure 5.2: Energy and entanglement spectrum of the Hofstadter model with $p/q = 3/7$ and two filled bands. (a): Details of energy spectrum showing lowest three energy bands. The Fermi level lies at $\mu = -1.9$ (gray horizontal line). The total Chern number of the two occupied bands is $C = 3$. Red (blue) vertical lines indicate $k_x$ at which the lower (upper) edge modes cross the Fermi energy ($k_x^L$ and $k_x^U$ in text). (b): Entanglement occupancies $f_a$. (c): Entanglement quasienergy. 20 levels are below the large ‘gap’ – the same as the number of occupied levels in the full system. Colored levels are also plotted in panels (b) and (d) with the same color scheme. (d): quasienergy in 2D polar coordinates. The radius is the quasienergy and the polar angle is $k_x$. The gray circle corresponds to $\gamma = 0$. There are three curves spiraling outward in the clockwise direction, corresponding to total Chern number $C = 3$. 
5.2.1 Recipe

To compute the entanglement spectrum at $k_x$, one first diagonalizes Eq. 5.5 and obtains its ground state correlation matrix, $\mathcal{G}_{yy'}(k_x) \equiv \langle c_y(k_x)\dagger c_{y'}(k_x) \rangle$. Hereafter we suppress the $k_x$ dependence. If $U$ diagonalizes $H$, viz., $U^\dagger HU = E \equiv \text{diag}(\varepsilon_1, \varepsilon_2, \cdots)$, then

$$\mathcal{G} = U f(E, \mu, T) U^\dagger, \quad (5.6)$$

where $f(\varepsilon, \mu, T) = \{\exp[(\varepsilon - \mu)/T] + 1\}^{-1}$ is the Fermi distribution. $\mathcal{G}$ is a $N_y \times N_y$ matrix. We will assume that $T = 0$, then $\mathcal{G}$ is the projection onto states with energy below the Fermi level $\mu$ (to be placed in a bulk spectral gap). The correlation matrix $G$ of subsystem $L$ is a submatrix of $\mathcal{G}$. Formally one can use a real-space projector $R$, which projects onto the lower subsystem, and is an oblong matrix of size $M \times N_y$. In the coordinate-space basis,

$$R = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 1 & 0 & \cdots & M \end{pmatrix}_{M \times N_y}, \quad (5.7)$$

thus

$$G = RGR^\dagger, \quad \text{Spec}(G) = \{f_1, f_2, \cdots, f_M\}. \quad (5.8)$$

$G$ is an $M \times M$ matrix, and its spectrum $\{f_a\}$ constitutes the single-particle entanglement spectrum (entanglement occupancy). Recall that the RDM has a form $\rho_L \propto e^{\hat{\Gamma}}$. The single-particle eigenvalues of $\hat{\Gamma}$ are known as the entanglement quasieenergies $\{\gamma_a\}$, and can be obtained by inverting the Fermi distribution

$$f_a = \frac{1}{\exp(\gamma_a) + 1}. \quad (5.9)$$

5.2.2 Counting the rank

The entanglement spectrum exhibits a rich structure. Let us first look at the broader strokes by counting the number of levels in three distinctive groups,
namely $f = 0$, $f = 1$, and $0 < f < 1$. Denote the number of zeros and ones as $D_0$ and $D_1$, respectively. They are by definition the dimensions of the kernels of $G$ and $I - G$. If $\nu$ bulk bands are occupied with periodic boundary conditions (hence no edge states interpolating the gaps), the total rank of $G(k_x)$ is $\nu N_y/q$, this is because each of the $q$ bands contains an equal number of states. Thus if $M \geq \nu N_y/q$, the rank of $G$ will also be $\nu N_y/q$. For $M \leq \nu N_y/q$, the rank of $G$ is $M$. Thus,

$$\text{rank}(G) = \min(M, \nu N_y/q) , \quad (5.10)$$

$$D_0 = M - \text{rank}(G) , \quad (5.11)$$

and similarly

$$\text{rank}(I - G) = \min\left\{M, N_y(1 - \nu/q)\right\} , \quad (5.12)$$

$$D_1 = M - \text{rank}(I - G) , \quad (5.13)$$

where $N_y(1 - \nu/q)$ is the rank of $I - G$, the projector onto unoccupied bands.

With cylindrical boundary conditions, the rank of $G$ changes with $k_x$ whenever an edge state crosses the Fermi level. As a result, $D_0$ and/or $D_1$ are discontinuous at such $k_x$ values for certain range of $M$. It is easy to verify that the condition for $D_0 = D_1 = 0$ is $M \leq \min(\nu/q, 1 - \nu/q) \times N_y$.

### 5.2.3 Characteristic features

We now turn to the finer structures in the entanglement spectrum. We use $p/q = 3/7$ flux per plaquette and open boundary conditions in $y$, and plot them in Fig. 5.1 (three filled bulk bands) and Fig. 5.2 (two filled bands). One can see that:

1. **Clustering at $f \sim 0$ and 1**: Most levels are clustered near $f = 0$ and $f = 1$. This reflects the fact that $E_F$ lies inside a bulk gap, forbidding existence of partially-occupied bulk states in the full system. Taking linear combinations of the occupied states in the full system, one can then create wavefunctions which are mostly confined to either the lower or the upper subsystem, yielding $f \sim 0$ and 1, respectively. We will have more to say about this when discussing Wannier functions.
2. **Spectral flow in \( f \):** A few levels interpolate between \( f \sim 0 \) and \( 1 \), clearly showing spectral flows as functions of \( k_x \), resembling that of the energy edge modes. The number of entanglement spectral flows is the same as the total Chern number \( C_{\text{occ}} \) of the filled bands, e.g. \( C_{\text{occ}} = 1 \) in Fig. 5.1 and \( C_{\text{occ}} = 3 \) in Fig. 5.2.

3. **Discontinuity in \( f \):** The occupancy \( f_a(k_x) \), for a fixed level index \( a \), is discontinuous at \( k_x = k_x^L \) where the lower edge state (the one localized near \( y = 1 \)) crosses the Fermi level (gray vertical lines in Fig. 5.1 and red vertical lines in Fig. 5.2). However, levels with consecutive \( a \) are connected at \( k_x^L \): \( f_a(k_x^L+0^+) = f_{a+1}(k_x^L-0^+) \).

There are as many such \( k_x^L \) values as the number of times the lower edge mode crosses the Fermi level, which is \( |C_{\text{occ}}| \). As one increases \( k_x \) through each such crossing, the number of levels with \( f \approx 1 \) and hence the total occupancy in the lower subsystem \( \sum_a f_a \) drops discontinuously by unity due to the exclusion of the edge mode. Eventually the \( f \approx 1 \) levels are repopulated due to the aforementioned spectral flow. In order to conserve the rank of \( G(k_x) \) (in the case of \( M > \text{rank}(G) \), cf. Eq. 5.11) upon increasing \( k_x \) by \( 2\pi \), then, there must be a discontinuous repopulation of the \( f \approx 0 \) levels. This occurs when the upper cylindrical boundary edge modes cross \( E_F \) at \( k_x = k_x^U \); these are the blue vertical lines in Fig. 5.2. As these modes have a vanishingly small projection onto the \( L \) subsystem in the thermodynamic limit, they lead to no discontinuity in the total occupancy of \( L \).

4. **Gap in quasienergy spectrum:** From the occupancy spectrum, one can invert the Fermi distribution (Eq. 5.9) to get the quasienergy spectrum \( \{\gamma_a\} \). A quasienergy plot more clearly reveals entanglement spectrum near \( f = 0 \) and \( f = 1 \), where many levels are clustered. In Fig. 5.2 (c), one can see that a substantial number of levels are clustered at \( \gamma \approx 140 \) (\( f \sim 10^{-60} \)) and are separated from the remaining levels by a pronounced gap. Actually this is a numerical artifact and these levels all lie at \( \gamma = \infty \). Recall the earlier result \( \text{rank}(G) = \min(M, \nu N_y/q) \) in Eq. 5.11 for a system with periodic boundary conditions. Here we have \( p/q = 3/7 \), \( N_y = 70 \), \( M = 35 \), and \( \nu = 2 \) since \( E_F \) is placed in the gap between the second and third bulk bands. Thus we would expect \( \text{rank}(G) = 20 \), and since the row dimension of \( G \) is \( M = 35 \), there should be 15 levels with \( f = 0 \), corresponding to \( \gamma = +\infty \). Had one looked at a system with more than half filling, one would find
entanglement quasienergies clustering at \( \gamma = -\infty \) instead, where the entanglement occupancy is exactly 1. These would correspond to the kernel of \( 1 - G \).

5. Discontinuity in quasienergy: Since our system has cylindrical boundary conditions, there are edge states, and there is a discontinuity in the quasienergy spectrum at each \( k^U_x \) and \( k^L_x \) value where lower and upper edge states are crossed by \( E_F \). When both edge states lie below \( E_F \), one counts 20 finite quasienergy levels. When \( k_x \) lies between consecutive \( k^U_x \) and \( k^L_x \) values, one of the lower boundary edge states has crossed the Fermi level, and the rank of \( G \) decreases to 19. The spectral flow in the vicinity of \( \gamma \approx 0 \) is continuous, however. Discontinuities in the entanglement energies occur for large values of \( |\gamma| \), where the occupancy is close to 0 of 1. When an edge state passes from below \( E_F \) to above \( E_F \), the rank of \( G \) changes discontinuously by \(-1\). For increasing \( k_x \), this occurs at any of the three \( k^L_x \) points in Fig. 5.2. Such an edge state has almost perfect projection onto the lower subsystem, hence its depopulation leads to a sudden rearrangement of entanglement levels with large negative quasienergies \( \gamma_a \) (\( f_a \approx 1 \)) and a loss of one such level. For \( k_x = k^U_x \), where the change \( \Delta \text{rank}(G) = +1 \), the ‘extra’ level enters via a discontinuous rearrangement of the levels with large positive quasienergies (\( f_a \approx 0 \)). (The situation is reversed if \( E_F \) lies within the first gap, in which case \( \Delta \text{rank}(G) = +1 \) at each \( k^L_x \) and \( \Delta \text{rank}(G) = -1 \) at each \( k^U_x \).) We see this clearly in Fig. 5.2, where the number of finite \( \gamma \) levels changes from 20 to 19 when \( k_x \) lies between consecutive \( k^L_x \) and \( k^U_x \) values.

5.2.4 Sensitivity to subsystem size

Since the magnetic unit cell is set along the \( y \)-direction, \( N_y \) and \( M \) must be chosen as integer multiples of \( q \) if there are to be an integer number of unit cells in the full system and/or the lower subsystem, respectively. As shown in Appendix B, changing \( N_y \) to \( N_y + m \) (\( m \in \mathbb{Z} \)) keeps the lower edge modes intact, but shifts the \( k_x \) values for the upper edge modes by \(-2\pi mp/q\). For example, the lower edge modes (red lines) are the same in Figs. 5.1 (a) and (b), but the upper edge modes (blue lines) in (b) are shifted in \( k_x \) by \(-6\pi/7\) relative to those in (a).

It turns out that changing \( M \) affects the entanglement occupancy in the
same way as changing $N_y$ would affect the upper edge modes. This is shown in the bottom panels of Figs. 5.1. Intuitively, this is because the entanglement cut serves as the upper edge for the lower subsystem. We should mention that keeping $M$ fixed while changing $N_y$ will not change the occupancy spectrum in any appreciable way because that only shifts the upper edge modes. Changing $N_y$ will thus change the $k_x^U$ values, and consequently where the rearrangements of the $f \approx 0$ parts of the entanglement spectrum occur, but will not affect the spectral flow for $\gamma \approx 0$. The reason will become more clear in the next section.

5.3 Adiabatic pumping of band projectors

The entanglement level occupancies $f_a$ are eigenvalues of the restricted correlation matrix $G = RGRT$. In searching for an intuitive picture of the various features of the entanglement spectrum, it is then natural to examine the unrestricted projector $G$. We found that much information can be extracted from $G$ itself.

5.3.1 $k_x - y$ covariance

In this section, we will use periodic boundary condition in the $y$-direction, i.e. $z = 1$ and $N_y \mod q = 0$ in Eq. 5.5. The Hamiltonian of Eq. 5.5 then satisfies

$$H(k_x + \phi) = T_y H(k_x) T_y^\dagger,$$

where $\phi = 2\pi p/q$ as before and where $T_y$ is the translation operator by one lattice spacing in the $y$-direction:

$$T_y = \begin{pmatrix} 1 & 0 \\ N_y-1 & 0 \end{pmatrix}.\quad (5.15)$$

The unitarity of $T_y$ guarantees that the bulk bands repeat themselves for $q$ times over the interval $k_x \in [0, 2\pi]$. With each successive increase of $k_x$ by $2\pi p/q$, the spectrum repeats and the corresponding energy eigenstates are shifted by $\Delta y = 1$. Denoting $B_j(k_x)$ as the projector onto the $j^{th}$ band, we have that $G(k_x) \equiv G_{\nu}(k_x)$
Figure 5.3: Full system band projectors for $p/q = 3/7$, $N_y = 28$, with periodic boundary conditions in $y$ and $k_x = 2\pi\kappa/q$. $B_j$ is the projector onto the $j$th band, and $G_\nu = B_1 + \cdots + B_\nu$ the projector onto the lowest $\nu$ bands. The magnitude of the matrix elements are represented by intensity and their sign by color (red positive, blue negative, white zero). $y_1$ and $y_2$ are the row and column indices of the projectors. Blue rules mark boundary of the magnetic unit cells. Red rules mark the bipartite cut, so the top-left quadrant of $G$ corresponds to the restricted correlation matrix $G$. Only $\kappa = 1$ and $3/2$ are shown here. Corresponding plots for other $\kappa$ values can be inferred from those shown here after shifting all matrix elements along the diagonal by $t$ as the solution of Eq. 5.21. Similarly, projectors of all half odd-integer $\kappa$ are obtained by shifting those of $\kappa = 3/2$. As $k_x$ is increased from 0 to $2\pi\kappa$ ($\kappa$ from 0 to $q$), the adiabatic pumping is evident in the diagonal motion of all matrix elements of the projectors. Note in particular that at integer $\kappa$, each diagonal block of $G_3$ consists of three sharply localized packets: the top-left one is contributed by $B_1$, whereas the other two result from the constructive addition of the diagonal blocks of $B_2$ and $B_3$. Similarly, at half odd-integer $\kappa$, $B_1$ and $B_2$ add constructively, yielding the two sharp packets in each diagonal block of $G_2$. The Chern number corresponds to how many magnetic unit cell boundaries (blue and red rules) any diagonal matrix element has passed by in one pumping period. Equivalently, it is the sum of diagonal matrix elements which are transferred across any magnetic unit cell boundary in one pumping period. For the $G_\nu$ type, it is intuitively how many packets are transferred.
is the projector onto the lowest \( \nu \) filled bands,

\[
G_\nu(k_x) = \sum_{j=1}^{\nu} B_j(k_x) .
\] (5.16)

The covariance in \( k_x \) and \( y \) is reflected as

\[
B_j(k_x + \phi; y_1, y_2) = B_j(k_x; y_1 + 1, y_2 + 1) ,
\] (5.17)

where \( y_1 \) and \( y_2 \) are row and column indices for \( B_j(k_x) \). Translational invariance on a scale of the magnetic unit cell corresponds to

\[
B_j(k_x; y_1, y_2) = B_j(k_x; y_1 + q, y_2 + q)
\] (5.18)

The same relations hold for \( G_\nu \).

### 5.3.2 Localized wavepackets in band projectors

While these projectors are explicitly constructed using the Bloch states, which are spatially extended, the fact that their eigenvalues are degenerate (either 0 or 1) means one may construct localized eigenstates around the cylinder, for each \( k_x \), by recombining Bloch states of different \( k_y \) with the same eigenvalue. In the continuum limit, where \( q \to \infty \) with \( p \) finite, these correspond to the familiar Landau strip basis. In fact, the projectors themselves are localized: An illustration is provided in Fig. 5.3, which shows several \( B_j \) and \( G_\nu \) for \( p/q = 3/7 \) at \( k_x = 2\pi\kappa/q \) for \( \kappa = 1 \) and \( \kappa = 3/2 \), both of which are local extrema of the energy bands.

The size of the magnetic unit cell naturally divides the projectors into blocks of size \( q \times q \). It is not surprising that the off-diagonal blocks drop exponentially, a consequence of the analyticity of \( B_j \) in complex \( k_y \)[16]. What is perhaps unexpected is that at band troughs (integer \( \kappa \) for odd \( j \) and half-odd-integer \( \kappa \) for even \( j \) in Fig. 5.3), the projectors \( B_j \), and especially \( G_\nu \), are quite well-localized even within the diagonal blocks. The diagonal matrix elements of the projectors correspond to electron density at the corresponding \( y \) coordinate. For single bands \( (B_j) \), any \( q \) consecutive diagonal elements sum to 1, thus one may think of them as constituting a wavepacket, and the projector \( B_j \) as consisting of \( N_y/q \) such wavepackets (one
per magnetic unit cell). For each wavepacket, the weight is dominated by one or two elements, as one can see in Fig. 5.3. The localization of the projector sums $G_\nu$ is even more prominent: when the gap between two neighboring bands is at a minimum, their projectors add constructively, resulting in two sharply localized dots on the diagonal line of $G_\nu$. For example, in Fig. 5.3, $B_2$ and $B_3$ add constructively at $\kappa = 1$, yielding the lower two dots in each diagonal block of $G_3$, and similarly, $B_1$ and $B_2$ add up to $G_2$ for $\kappa = 3/2$. Now, $q$ consecutive diagonal elements in $G_\nu$ must sum to $\nu$, thus each of the $\nu$ dots can be intuitively understood as one localized wavepacket. The constructive superposition of neighboring bands then indicates the corresponding single-band wavepackets have opposite parity so that the off-diagonal elements cancel each other. To relate to the aforementioned “strip” states, we note that any column of a projector is an eigenstate of the same projector, with eigenvalue $1$. The diagonal nature of these band projectors thus ensures the existence of such strip states.

5.3.3 Adiabatic pumping of wavepackets and the Diophantine equation

We now establish a connection between the band projectors and the seminal work of Thouless et al. [64] on the Chern numbers for the Hofstadter bands. Consider first the individual band projectors $B_j$. We write

$$y = q\ell + m , \quad k_x = \frac{2\pi\kappa}{q} ,$$

(5.19)

where $\ell$ and $m$ are integers. Thus $\ell$ is the magnetic unit cell coordinate, and $m$ the coordinate within each such cell. For a single band, denote the position of any of its wavepackets as $m(\kappa)$, then $(k_x, y)$ covariance implies

$$m(\kappa + tp) = m(\kappa) - t , \quad t \in \mathbb{Z} .$$

(5.20)

$^1$Since any projector squares to itself, each column is an eigenstate with eigenvalue 1. However, they are not normalized, nor are they orthogonal to each other (because no eigenstate with eigenvalue 0 is present). In fact the construction of Wannier states is one way to orthonormalize these eigenstates.
Of course both $m$ and $\kappa$ are only defined modulo $q$. The relevant quantity in the $k_x$ pumping is the ‘velocity’ of the packet (with $k_x$ as ‘time’), *i.e.*, the number of sites it traverses when $\kappa$ is *effectively* increased by 1:

$$tp = sq + 1 \quad , \quad s \in \mathbb{Z} \; , \; |t| < q \; . \quad (5.21)$$

A graphical construction is shown in Fig. 5.4. Clearly, $t$ will be the number of packets transported through any given boundary during the cycle $k_x \rightarrow k_x + 2\pi$ (number of blue flow lines in the figure). It is also equal to the number of magnetic cells traversed by a single packet. There is however a mod $q$ ambiguity associated with the sign indeterminancy of $t$, *e.g.* for $p/q = 3/7$, one has that $(t, s) = (5, 2)$ and $(-2, -1)$ both satisfy Eq. 5.21. In general, without looking at intermediate $\kappa$ snapshots, one cannot tell if the packet had advanced by $t$ or retreated by $q - t$. We have examined different $p/q$ ratios on both square and triangular lattices, and we find that for the lowest band on a square lattice, the ambiguity can always be resolved, without needing to inspect intermediate $\kappa$, by picking the value of $t$ which has the smaller magnitude $|t|$, *e.g.* $t = -2$ instead of 5 for $B_1$ in Fig. 5.3. Intuitively, this means the packet moves toward the nearest possible position allowed by Eq. 5.20.

Eq. 5.21 is recognized as the Diophantine equation of TKNN[64] with $r = 1$, according to which $t$ is simply $C_1$, the Chern number of the lowest band. The heuristic of taking the smaller $|C_1|$ in resolving the mod $q$ ambiguity agrees with ref. [64]. Since this picture does *not* distinguish between different bands, the Chern numbers of all bands are equivalent mod $q$.

The correlation matrix $G_\nu$ for $\nu$ occupied bands has $\nu$ packets per diagonal block, with each moving according to Eq. 5.21, as required by the $(k_x, y)$ covariance. However, their *collective* motion depends on their relative spacing. Consider for example the diagonal blocks of $G_3$ for $\kappa = 1$, as shown in top right panel of Fig. 5.3. In each diagonal block, the diagonal elements with highest weight (packets) are at $y = (2, 4, 7)$. Then $(k_x, y)$ covariance requires that at $\kappa = 2$, they are moved to $y = (2 - t, 4 - t, 7 - t) = (4, 6, 9)$ where $t = -2$ is the solution to Eq. 5.21 as discussed before. Since $y = 9$ is simply the $y = 2$ element of the next diagonal block, thus in each diagonal block, the packets of $\kappa = 2$ are at
Figure 5.4: \((m, \kappa)\) construction for \(p/q = 3/7\), assuming \(m = 0\) when \(\kappa = 0\). Red line: flow in the order of \(T_y\) translation whereby \(\kappa \rightarrow \kappa + p\) and \(y \rightarrow y - 1\). Blue lines: flow in the order of \(k_x\) pumping. The number of times a packet is transferred across any boundary line, \(e.g.\ m = 0\) line, is the same as the number of blue flow lines crossing the boundary line. There is a mod \(q\) ambiguity as can be seen from the validity of both the solid and dashed blue flows.
\( y = (4, 6, 9) = (4, 6, 2) \mod 7 \). Comparing this with those of \( \kappa = 1 \), one can see that, effectively, only one packet moved from \( y = 7 \) to \( y = 6 \).

We found that for any \( \nu \neq q \), this observation holds true (for \( \nu = q \), \( \mathcal{G}_\nu \) is the identity). That is to say, as \( \kappa \to \kappa + 1 \), the net effect is for only one of the \( \nu \) packets to change position. We will not attempt to explain this observation, but rather take it as a starting point, and explore its implications. Mathematically, this observation – the reduction of the motion of multiple wavepackets to that of a single mobile packet at a time – means that the positions of dominant diagonal elements (the packets) can be labeled in such a way that

\[
    m_i(\kappa + 1) = m_{i+1}(\kappa) \quad i = 1, 2, \ldots, \nu - 1
\]

with each \( m_i \) still satisfying Eq. 5.20, i.e. \( m_i(\kappa + tp) = m_i(\kappa) - t \). Note that this does not mean \( m_i \) can be identified with the wavepacket of a single band: they result from constructive superposition of single-band projectors, as discussed earlier.

To illustrate Eq. 5.22, take again \( \mathcal{G}_3 \) as an example: at \( \kappa = 1 \), \( (m_1, m_2, m_3) = (7, 2, 4) \), while at \( \kappa = 2 \), \( (m_1, m_2, m_3) = (9, 4, 6) = (2, 4, 6) \mod 7 \). Then as \( \kappa \) increases by 1, the effective change is of one packet (the mobile one) moving from \( m_1(\kappa) \) to \( m_\nu(\kappa + 1) \) with stride \( t_\nu \),

\[
    m_\nu(\kappa + 1) = m_1(\kappa) - t_\nu \quad |t_\nu| < q
\]

The RHS is therefore \( m_1(\kappa + t_\nu p) \), while the LHS is

\[
    m_\nu(\kappa + 1) = m_{\nu-1}(\kappa + 2) = \cdots = m_1(\kappa + \nu)
\]

from Eq. 5.22. Thus \( t_\nu \) is determined by

\[
    t_\nu p = sq + \nu \quad s \in \mathbb{Z}, \quad |t_\nu| < q
\]

Again, there is a mod \( q \) ambiguity because of the sign indeterminacy of \( t_\nu \). For square lattice, the heuristic of using the smaller \( |t_\nu| \) still seems to hold, e.g. while both \( (t_3, s) = (1, 0) \) and \( (t_3, s) = (-6, -3) \) satisfy Eq. 5.25, the actual system picks \( t_3 = 1 \). Ref. [65] mentioned that \( s \) and \( t_\nu \) cannot simultaneously be odd for
either the hexagonal or triangular lattices. Incidentally, for \( \nu = p \), \( t_p = 1 \) is always a solution with the corresponding \( s = 0 \), i.e. the total Chern number of the lowest \( p \) bands is always 1.

Eq. 5.25 is the TKNN Diophantine equation[64] for \( r = \nu \). There, \( t_\nu \) is the total Hall conductivity (the sum of the Chern numbers) of the \( \nu \) occupied bands. It is also the winding number of the energy edge states in the \( \nu^{th} \) gap[27]. We now have a third interpretation: it is the number of sites traversed by the mobile packet during each \( \kappa \) increment. Equivalently, it is the number of mobile packets transported across any magnetic unit cell boundary during the cycle \( k_x \rightarrow k_x + 2\pi \).

The entanglement spectrum can now be understood intuitively. Whenever the mobile packet leaves the lower half-cylinder through the cut between \( M \) and \( M + 1 \) (in Fig. 5.3, proceeding from top-left quadrant through the red line into the lower-right quadrant), there is an occupancy flow from \( f = 1 \) to 0. The number of flow lines is then equal to the number of packets which move through the cut, which is the total Chern number. In the periodic \( y \) boundary case, where the cylinder is compactified into a torus, the flow across \( M = N_y/2 \), is always concomitant with another packet moving from \( y = N_y \) to \( y = 1 \), hence a symmetric flow from 0 to 1 with its wavefunction localized at the opposite end. (Entanglement occupancy with periodic \( y \) boundary is shown in Fig. 5.5 in the next section). Furthermore, if we change the position of the entanglement cut \( M \) (not necessarily along a magnetic cell boundary, for example), this will simply change the value of \( k_x \) when a packet hits the cut, whence the \( k_x \) translation shown in Figs. 5.1

While the entanglement spectrum only reveals the total Chern number, the correlation matrix retains some information about the individual Chern numbers of constituent bands, manifested as the separation between its wavepackets. Note that Eq. 5.23 can be taken as a definition of \( t_\nu \) with arbitrary \( \nu < q \), without interpreting \( \nu \) as the number of filled bands. After all, the total Chern number of the lowest \( \nu \) bands is the same whether or not they are filled. We explicitly replace \( \nu \) with \( n < q \) below to avoid any such connotation. From Eq. 5.22 and 5.23, we have

\[
m_{n+1}(\kappa) - m_n(\kappa) = t_{n-1} - t_n = -C_n ,
\]  
(5.26)
thus the two packets at $m_{n+1}$ and $m_n$ are separated by a distance of $C_n$. Quantities such as the four-point correlation $\mathcal{F}(\Delta) = \langle c_y^\dagger c_y c_y^\dagger c_{y+\Delta} \rangle$ thus have peaks at $\Delta = C_i$ apart from $\Delta = q, 2q, \text{etc.}$.

On a square lattice, the Hofstadter model exhibits a particle-hole symmetry. This implies that $C_j = C_{q+1-j}$. For even $q$, the bulk spectrum is known to have no central gap[74, 41], therefore the Chern numbers of the two central bands are not individually well defined, and one can speak only of a Chern number for the pair. It is interesting to notice its implication on the distribution of the wavepackets within each unit cell: if on the contrary there is a central gap, then $t_{q/2} = C_1 + C_2 + \cdots + C_{q/2} = C_q + C_{q-1} + \cdots C_{q/2+1}$. Since the total Chern number of all bands must be zero, we must have $t_{q/2} = 0$. Now according to Eq. 5.22 and 5.23, $m_{q/2+1}(\kappa) = m_1(\kappa) - t_{q/2} = m_1(\kappa)$, so the $(q/2 + 1)^\text{th}$ packet and the first one are forced onto the same site. Thus, the fact that there is no central gap in this case guarantees that there will be no packet ‘collisions’.

The natural question to ask next is how the wave packets are arranged when $\nu > q/2$. To illustrate this, consider a specific case with $p/q = 5/8$ and $\nu = 7$ filled bands. The Chern numbers of the lowest three bands are $-3$, $5$, and $-3$, respectively, and particle-hole symmetry guarantees that these values repeat for the upper three bands. The central two bands therefore have a combined Chern number of $C_{4,5} = 2$. Since Chern numbers represent the separation between wave packets, we can fill in the first four packets with no difficulty (the position of the first packet being arbitrary). The location of the fifth packet cannot be determined because $C_4$ is not well-defined, but the location of the sixth packet is found by shifting the fourth one by $C_{4,5} = 2$. The rest of the packets can be filled in a similar fashion. Thus the vanishing of the central gap implies an indeterminacy of the position of the $(q/2 + 1)^\text{th}$ wavepacket. To resolve this, one can add in an infinitesimal second-neighbor hopping that breaks the particle-hole symmetry and results in a small central gap. For example, one can introduce a second-neighbor hopping $t'$ along one of the two diagonals in each unit cell (say in the direction $\hat{x} - \hat{y}$). This construction interpolates between the square lattice when $t' = 0$, and the triangular lattice when $t' = 1$ (see Appendix C). For $t' \ll 1$, we find that the
Chern numbers of the central two bands are resolved as $C_4 = 5$ and $C_5 = -3$. $C_4$ can now be used to determine the position of the fifth packet.

### 5.4 Wannier center flow

The Diophantine equation (5.25) describes a mod-$q$ property of the Hofstadter problem, which is a result of the $(k_x, y)$ covariance of Eq. 5.17. No knowledge of intermediate values of $k_x \in [\kappa \phi, (\kappa + 1) \phi]$ is required in obtaining Eq. 5.25. This comes at a price of the ambiguity in $t \pmod{q}$ and $s \pmod{p}$, which intuitively contain the information of the direction in which any given packet is moving. In this section, we settle this issue by examining the localized eigenstates of the projectors over the full range of $k_x$, i.e. the Wannier functions.

#### 5.4.1 Wannier functions in 1D

The application of Wannier functions to the analysis of topological band structures has recently been developed in refs. [62], [80], and [52]. Following these references, consider first a periodic one-dimensional system consisting of $N$ unit cells with $q$ internal degrees of freedom per cell. Let $X$ be a cell coordinate and let $m$ index the internal degree of freedom; we may take $X \in \{1, \ldots, N\}$ and $m \in \{1, \ldots, q\}$. Bloch’s theorem says $\Psi_{n,k}(X,m) = e^{ikX} u_{n,k}(m)$, where $n$ labels the $q$ bands. One may thus decompose the Hilbert space as $\mathcal{H} = \mathcal{H}_X \otimes \mathcal{H}_m$, writing $|\Psi_{n,k}\rangle = |k\rangle \otimes |u_{n,k}\rangle$. Here $|u_{n,k}\rangle$ is an eigenstate of the Fourier transform $H_{mm'}(k) \equiv \langle X,m | \hat{H} | X', m' \rangle$, i.e. it is a Bloch cell function.

For a system with periodic boundary conditions, the position operator can be taken to be $U = e^{2\pi i X/N}$, as in the work of Yu et al.[80] An eigenstate of $\tilde{U} \equiv PUP$, where $P$ is a projector onto a subset of energy bands, is of the form

$$|\Phi_\lambda\rangle = \sum_{n,k} \Phi_{n,k}^\lambda |k\rangle \otimes |u_{n,k}\rangle,$$

where the sum on the band index $n$ is over the desired subset, and $\lambda$ labels the eigenvalues. Demanding $\tilde{U} |\Phi_\lambda\rangle = e^{2\pi i \lambda/N} |\Phi_\lambda\rangle$, one obtains $\Phi_{m,k+\Delta k}^\lambda = e^{-2\pi i \lambda/N} M_{mn}(k+\Delta k) |u_{n',k}\rangle$.
\( \frac{1}{2} \Delta k \) \( \Phi_{n,k}^{\lambda} \) (sum on \( n \) over selected bands), where

\[
M_{mn}(k) = \langle u_{m,k+\frac{1}{2} \Delta k} | u_{n,k-\frac{1}{2} \Delta k} \rangle ,
\]

with \( \Delta k = \frac{2 \pi}{N} \). The eigenvalue equation, which follows from setting \( \Phi_{n,0}^{\lambda} = \Phi_{n,2\pi}^{\lambda} \), is then

\[
det(e^{2\pi i \lambda} - W) = 0 ,
\]

where \( W = M(N \Delta k - \frac{1}{2} \Delta k) \cdots M(\Delta k - \frac{1}{2} \Delta k) \) is a Wilson loop. Note that \( \lambda \) is not necessarily real since \( \tilde{U} \) is the projection of a unitary operator but is not unitary itself. In a more general setting, where the wavefunctions \( u_m \) depend on a set of parameters \( g \), one has

\[
\langle u_m(g + \frac{1}{2} \Delta g) | u_n(g - \frac{1}{2} \Delta g) \rangle = \left[ \exp \left( i A^\mu \Delta g_\mu - \frac{1}{2} Q^{\mu\nu} \Delta g_\mu \Delta g_\nu + \mathcal{O}(\Delta g^3) \right) \right]_{mn},
\]

where \( A^\mu \) is the nonabelian Berry connection,

\[
A^{\mu}_{mn}(g) = i \langle u_m | \frac{\partial u_n}{\partial g_\mu} \rangle
\]

and \( Q^{\mu\nu} \) is the quantum geometric tensor\[51, 13\],

\[
Q^{\mu\nu}_{mn}(g) = \left( \frac{\partial u_m}{\partial g_\mu} \right) (1 - P) \left( \frac{\partial u_n}{\partial g_\nu} \right).
\]

In our case, as \( N \to \infty \) we have that the Wilson loop becomes unitary, and each eigenvalue \( \lambda \) is real.

For a single band, we can write

\[
\lambda_I = \int_0^{2\pi} \frac{dk}{2\pi} A(k) + I ,
\]

where \( I \) is an integer and \( A(k) = i \langle u(k) | \frac{\partial}{\partial k} | u(k) \rangle \). Thus for a single band, the state \( |\Phi_{\lambda_I}\rangle \) is localized at unit cell \( I \) with an offset \( \gamma/2\pi = \int_0^{2\pi} dk A(k) \).

When the internal space of \( |u_{n,k}\rangle \) coincides with real space (\( e.g. \) the lattice site \( m \) within the magnetic unit cell in Hofstadter problem), one may refine the definition of the position operator, writing

\[
U \to e^{2\pi i \hat{X}/N} e^{2\pi i \hat{m}/qN} ,
\]
where $\hat{m} = \sum_{m=1}^{q} m \ket{m}\bra{m}$ measures the position within each unit cell. For the single band case, this shifts the offsets $\gamma$ to

$$\tilde{\gamma} = \gamma + q^{-1} \int_{0}^{2\pi} dk \bra{u_{n,k}} \hat{m} \ket{u_{n,k}} .$$

(5.35)

Equivalently, one may also introduce the modified cell functions $\ket{\tilde{u}_{n,k}}$,

$$\ket{\tilde{u}_{n,k}} = q^{-1/2} \sum_{m=1}^{q} \bra{m} u_{n,k} \rangle e^{\frac{-ikm}{q}} \ket{m}$$

(5.36)

and use them in computing the Berry connection.

When the projector $P$ is onto multiple bands, the Wilson loop becomes a matrix, and the eigenvalues of $\tilde{U}$ are $e^{2\pi i \lambda_{I,w}/N}$, where $w$ is an additional label running from one to the number of bands, i.e. the dimension of the projector[52]. One then has

$$\lambda_{w,I} = I + \frac{\theta_{w}}{2\pi} .$$

(5.37)

Again, for systems where the internal ‘orbital space’ corresponds to real space, one can refine the position operator as in Eq. 5.34.

For two-dimensional lattices, Wannier functions can be defined at each $k_x$. For a single band $n$, its $y$-center is the Berry phase $\gamma_n(k_x)/2\pi$. The band Chern number is the negative of the winding number of $\gamma_n(k_x)$ over the interval $k_x \in [0, 2\pi]$:

$$C_n = \frac{1}{2\pi i} \int_{\text{BZ}} d^2k \nabla_k \times \bra{\psi_n(k)} \nabla_k \ket{\psi_n(k)} \cdot \hat{z}$$

$$= \frac{\gamma_n(0) - \gamma_n(2\pi)}{2\pi} .$$

(5.38)

Thus the Wannier center shifts by $-C_n$ (magnetic) unit cells over $k_x \rightarrow k_x + 2\pi$, as found by Qi in ref. [52]. We have seen in §5.3 that the packet associated with band $n$ is translated by $-C_n$ lattice sites during $k_x \rightarrow k_x + 2\pi/q$, so over $k_x \rightarrow k_x + 2\pi$, it will be translated by $-C_n$ magnetic unit cells, in agreement with the Wannier picture.

For multiple bands, we have

$$\det W = \exp \left\{ i \int_{0}^{2\pi} dk \tr A(k) \right\}$$

(5.39)
hence
\[ \sum_{w=1}^{\nu} \theta_w = \sum_{n=1}^{\nu} \gamma_n , \quad (5.40) \]
where \( \nu = \text{rank}(P) \). Then similar to the single band case, we conclude that in 2D, the total shift of all (inequivalent) Wannier centers is given by the sum of the Chern numbers. This is reflected in §5.3 as the number of mobile packets transported through any given magnetic unit cell boundary.

When there is no level crossing among \( \{ \lambda_{w,I} \} \) over the period of \( k_x \), one can combine the \( w \) and \( I \) indices. Define a composite index \( \mu(w,I) = w + \nu I \), with \( \lambda_{w,I} \rightarrow \lambda_{\mu} \). Sending \( k_x \) from 0 to \( 2\pi \) amounts to an index shift, which is universal for all \( \theta_\mu \) since there is no level crossing. Then the eigenfunctions \( \Theta_\mu(k_x) \) of \( W \) (here taken to be periodic in the index \( \text{modulo} \ q \)) satisfy
\[ \Theta_\mu(2\pi) = \Theta_{\mu+\sigma}(0) , \quad \sigma \in \mathbb{Z} \quad (5.41) \]
which is just a cyclic permutation in the \( w \) index with an offset \( \sigma \). One may think of the set of \( \{ e^{i\theta_w} \} \) as \( \nu \) points on the unit circle where different indices \( I \) are equivalent. Then during the \( \sigma \) cyclic permutation, the perimeter of the circle is covered by these \( \nu \) points \( \sigma \) times, i.e.,
\[ \frac{1}{2\pi} \sum_{w=1}^{\nu} [\theta_{w,I}(2\pi) - \theta_{w,I}(0)] = \sigma . \quad (5.42) \]
But according to Eq. 5.40, the LHS is simply the total Chern number of constituent bands,
\[ \sigma = - \sum_{n=1}^{\nu} C_n . \quad (5.43) \]

### 5.4.2 Wannier center flow in Hofstadter problem and a general relation with the entanglement spectrum

For the Hofstadter model, we have numerically diagonalized the operator \( \tilde{U} = P e^{2\pi i Y/N_y} e^{2\pi i m/q N_y} P \), where \( Y \in \{1, \ldots, N_y\} \) runs over the unit cells, and \( m \in \{1, \ldots, q\} \) runs over the individual sites within each unit cell. States with zero
Figure 5.5: Wannier centers vs. entanglement occupancy for \( p/q = 3/7 \) with different filling fraction and 4 magnetic unit cells in the full system and 2 in the half system. Top: Wannier centers using lattice coordinate \( \hat{y} \) as position operator. Colored dots: \( \theta_y \). Colored lines: \( \langle y \rangle_\theta \). Black-white background: diagonal elements of the lowest band projector \( B_1 \), black = 1, white = 0. See also Fig. 5.3. Center: Wannier center using \( \langle Y \rangle_\theta \) (magnetic unit cell coordinate). Corresponding levels have the same color as in top panels. Bottom: entanglement occupancy (no type/color coding). In (a), each type/color of point in the top and center panels corresponds to a packet, i.e., related via \( (k_x, y) \) covariance, in this case advancing by \( p = 3 \) in \( \kappa \) as \( y \rightarrow y + 6 \) (equivalent to \( y \rightarrow y - 1 \)). In (b) and (c), each type/color of point corresponds to a mobile packet (i.e. not related by the \( (k_x, y) \) covariance). It is also clear that within each magnetic cell, only one mobile packet moves as \( \kappa \rightarrow \kappa + 1 \), with \( \kappa = \) half-odd-integer in (b) and integer in (c). Notice the similarity between the Wannier center flow and the entanglement flow: a packet (a) or mobile packet (b and c) crossing the magnetic cell within the bulk has almost the same shape as the entanglement downflow; while crossing from \( y = N_y \) to \( y = 1 \) has almost the same shape as the entanglement upflow. The plateau-like feature in the entanglement flow lines can be traced back to the \( y \) plot in the top panels as the motion of a (mobile) packet within one magnetic cell.
Figure 5.6: Normalized eigenfunctions of $GRG$ ($\psi_a(y)$ in text) for $p/q = 3/7$, $N_y = 112$, $M = 56$, and $\nu = 3$ at $k_x = 2\pi/7$ (cylindrical boundary conditions). The Chern number of the filled bands is $C_{\text{occ}} = 1$. Color corresponds to the sign of the wavefunction (red for positive, blue for negative), and intensity to amplitude (white for zero intensity).
eigenvalues are those projected out by $P$. For a single band, $P = B_n$, while $P = \mathcal{G}_\nu$ for $\nu$ filled bands. We write $\tilde{U} |\theta\rangle = e^{i\theta} |\theta\rangle$, and we compute three slightly different Wannier centers: $y_\theta \equiv \theta N_y/2\pi$, $\langle Y \rangle_\theta \equiv \langle \theta | \hat{Y} | \theta \rangle$, and $\langle y \rangle_\theta \equiv \langle \theta | \hat{Y} + \frac{m}{q} | \theta \rangle$, where $\theta$ is wrapped in such a way that $y_\theta$ is restricted to $[0.5, N_y + 0.5]$. The top panel of fig. 5.5a shows the Wannier centers defined above for the lowest band of $p/q = 3/7$ with $N_y = 28$. The $y_\theta$ values, shown as colored dots, have a proper translational property: these values are generated by shifting any single flow by successive multiples of $q$. In the vicinity of half-odd-integer $\kappa$, a Wannier center migrates from one site to the next one which is $-C_1 = 2$ sites ahead. In terms of the corresponding wavefunction (not plotted), what happens is that around integer $\kappa$, it has a single peak at the site given by its (rounded) eigenvalue $y_\theta$. As $\kappa$ slowly moves toward the next half-odd-integer value, some weight is transferred to the next site, causing the eigenvalue $y_\theta$ to interpolate between the two values. As $k_x$ is increased by $2\pi$, $q$ such migrations are made, i.e., each Wannier center is shifted backwards by $C_1$ magnetic unit cells. In the bulk, $\langle y \rangle_\theta$ (colored lines in (a)) and $y_\theta$ overlap. Near $y = N_y$, part of the weight is pushed over the boundary to the $y = 1$ end, thus $\langle y \rangle_\theta$ starts to deviate from $y_\theta$ and drops, until all weight is transferred to the other boundary.

If instead of $\tilde{U}$, we were to diagonalize the operator $\tilde{V} \equiv P \left( \frac{\hat{Y}}{N_y} + \frac{m}{q N_y} \right) P$, then the behavior in the bulk will be the same, but near the edge there will be avoided crossings in the Wannier center flow of $\langle y \rangle_\theta$. The flow of the magnetic unit cell coordinate $\langle Y \rangle_\theta$, shown as colored lines in fig. 5.5b, is similar to that of $\langle y \rangle_\theta$, but with an emphasis on the occasions when $\langle y \rangle_\theta$ crosses a magnetic cell boundary.

The Wannier centers of two and three filled bands are shown in Fig. 5.5. As $\kappa$ increases by 1, only one Wannier center per magnetic unit cell flows by $\sigma = -\sum_{i=1}^\nu C_i$ sites (Eq. 5.43), corresponding to the motion of a mobile packet in §5.3. $\sigma = -3$ for two filled bands and 1 for three filled bands. There is no level crossing in $y_\theta$, hence Eq. 5.43 holds true: following any flow line from $k_x = 0$ to $2\pi$ leads one to $-\sigma$ levels beneath the starting point. Notice the same behavior in the entanglement quasienergy of Fig. 5.2.

There is a striking similarity between the magnetic cell coordinate flow $\langle Y \rangle$
Figure 5.7: Normalized eigenfunctions of $RGR$ ($\tilde{\psi}_a(y)$ in text) for the parameters given in the caption to Fig. 5.6. The entanglement energies are plotted in the top half. Grey points correspond to $\gamma_a = \infty$ ($f_a = 0$). As expected, states with entanglement energy $\gamma_a \approx 0$ are localized in the vicinity of the cut.
and the entanglement occupancy spectrum, shown in the center and bottom rows of Fig. 5.5: the upward flow of $\langle Y \rangle$ looks exactly like the upward flow of $f$, while the downward flow of $\langle Y \rangle$ within each $Y \rightarrow Y - 1$ sector looks like the downward flow of $f$. This can be understood in the following way: the spectrum of $\hat{Y}$ can be thought of as a coarse-grained version of $\hat{y}$, and should look just like $\langle Y \rangle$ and $\langle y \rangle$ shown in Fig. 5.5, so the effect of coarse-graining is to suppress the flow within a coarse-grained cell, and enhance the flow migrating between different cells. One may pushing the coarse graining to the extreme where all sites with $y \leq M$ count as $\bar{y} = 0$, and all above $M$ as $\bar{y} = 1$. Then the only significant flow is from $\bar{y} = 0$ to 1, corresponding to $Y = 2$ to $Y = 3$ in Fig. 5.5, and from $\bar{y} = 1$ to $\bar{y} = 0$, corresponding to $Y = 4$ through the periodic boundary to $Y = 1$. In Fig. 5.5, such a coarse graining would keep the upward $Y$ flow intact, while push $Y = 3$ and $Y = 2$ lines to top and bottom respectively for the downward flow, making it look just like the entanglement occupancy spectrum. In fact, one can prove that the coarse-grained Wannier spectrum is identical to the entanglement spectrum: Consider two arbitrary projectors $P$ and $R$. One can think of them as two matrices of the same dimension (zero-padded, if necessary, to fill out the dimensions). We claim that $PRP$ and $RPR$ have identical eigenspectra. To see this, assume $|\psi_a\rangle$ is an eigenstate of $PRP$ with non-zero eigenvalue $\lambda_a$. Then

$$|\psi_a\rangle = \frac{1}{\lambda_a} PRP |\psi_a\rangle$$

and therefore $P|\psi_a\rangle = |\psi_a\rangle$. Thus, $|\psi_a\rangle$ is an eigenstate of $P$ with eigenvalue 1. It then follows that

$$RPR |\psi_a\rangle = \lambda_a R |\psi_a\rangle$$

from which it follows (using $R^2 = R$) that $|\tilde{\psi}_a\rangle = R |\psi_a\rangle$ is an eigenstate of $RPR$ with the same eigenvalue $\lambda_a$. Thus the non-zero spectrum of $PRP$ belongs in that of $RPR$, and vice versa, so they are identical. Since the coarse-grained position operator $R$ is (the complement of) the projector used in constructing the restricted correlation matrix $G$, the entanglement spectrum of $G = RGR$ is identical to the coarse-grained Wannier centers $GRG$, and the entanglement eigenstates are
obtained by projecting the coarse-grained Wannier states onto the relevant half space.

In Fig. 5.6 we plot the normalized eigenfunctions $\psi_a(y)$ of $GRG$ for the case $p/q = 3/7$, $N_y = 112$, $M = 56$, and $\nu = 3$ at $k_x = 2\pi/7$. Note how the behavior of the eigenfunctions mimics that of the Wannier states, with $\psi_a(y)$ localized at a position which moves across the entire cylinder as the label $a$ advances through a range corresponding to the rank of $GRG$ ($3 \times 112 = 48$ in this case; see §5.2.2). States #1 through #64 belong to the kernel of $GRG$ and are all degenerate. The insertion of the real space projector $R$ thus fails to resolve these wavefunctions in real space, which explains the speckled pattern on the left side of the figure. The Wannier states are better localized however, since $R$ may be considered a coarse-grained approximation to $y$.

In Fig. 5.7, we plot the normalized eigenfunctions $\tilde{\psi}_a(y)$ of $RGR$ for the same parameters, along with the corresponding entanglement energies $\gamma_a$. According to our definitions,

$$|\tilde{\psi}_a\rangle = R |\psi_a\rangle / \sqrt{\langle \psi_a | R | \psi_a \rangle}.$$  \hspace{1cm} (5.46)

Note how states of large positive $\gamma_a$ ($f_a \approx 0$) as well as states of large negative $\gamma_a$ ($f_a \approx 1$) are localized far from the $A/B$ boundary.

### 5.5 Summary

In this chapter, we have studied the entanglement spectrum and Wannier center flows of the Hofstadter problem. Most of the data presented in this chapter was for the square lattice with $p/q = 3/7$ flux quanta per unit cell, but most of our observations are robust with respect to changing lattices, fluxes, and fillings. The entanglement spectrum of a subsystem exhibits spectral flow similar to that of the full system’s energy edge modes: the total Chern number controls the number of flow lines, and its sign tells the direction of the flow. When cylindrical boundary conditions are used in the full system, the entanglement spectrum exhibits level index discontinuity on the flow line. This is a manifestation of the crossing of the Fermi energy with the full system edge modes, which results in a total occupancy
The behavior of the entanglement spectrum can be understood by looking at the full system band projectors. These projectors are well localized and thus represented by wavepackets on their diagonals. For single bands, the packets flow under $k_x$ pumping. The $(k_x, y)$ covariance then imposes restrictions on possible flow rate, described by a Diophantine equation first derived by TKNN[64]. Since the Chern numbers are also given by the same equation, the topological property of the system can be described equivalently in terms of the motion of these packets: the number of magnetic unit cells traversed by each packet during one period of $k_x$ is given by the Chern number, with the direction given by its sign. For multiple bands, the flow is that of the mobile packets moving under $k_x$ pumping, and the number of mobile packets crossing a given boundary gives the total Chern number of filled bands. The entanglement spectrum can then be understood as a measure of detecting when these packets cross a particular boundary, namely the entanglement cut.

Using the $(k_x, y)$ covariance alone (and hence the Diophantine equations) only fixes the flow and the Chern numbers up to mod $q$ because it only relates different $k_x$ points of fixed separation of $2\pi/q$. The localization of the projectors suggests the use of Wannier functions for smooth interpolation between these $k_x$ points. For single bands, the Wannier center at each $k_x$ is given by the corresponding Berry phase and is represented by one packet in the projector diagonal. The flow of the Wannier center is then described by the winding number of this Berry phase, which is the band Chern number. For multiple bands, the Berry phase is replaced by a set of eigenvalues of the Wilson loop operator. If there is no level crossing over the full range of $k_x$, then all levels experience a universal index bump of $\sigma$ as $k_x \rightarrow k_x + 2\pi$, and $\sigma$ is given by the sum of Chern numbers. In computing the Wannier center, the position operator can be either the (magnetic) unit cell coordinate alone, or one that also includes the internal coordinates (lattice cell within each magnetic cell). The spectrum of the former is a coarse-grained version of the
latter. One can take the coarse graining to the extreme of a bipartition, at which point the position operator becomes a real space projector, and the coarse-grained Wannier spectrum becomes identical to the entanglement occupancy spectrum.

Chapter 6

Topological indices for open and thermal systems

The main tools in topological classification are Berry phase in 1D and Chern number or its variants in 2D and 3D. However, their definition relies on the notion of wavefunctions. For systems described not by wavefunctions but density matrices, there is no natural topological characterization through these concepts. For example, a thermal averaged Chern number includes contributions from bands unfilled at zero temperature and is therefore no longer quantized. In this chapter, we investigate the possibility of defining quantized topological indices for 2D systems described by density matrices, with the desiderata that it reproduce the pure state topological indices at zero temperature, and become trivial at infinite temperature.

Recall that the Chern number of a 2D system can be formulated as the winding number \( C = [\gamma(2\pi) - \gamma(0)]/2\pi \), where \( \gamma(k_x) \) is the Berry phase, computed at fixed Bloch wave vector \( k_x \), along the path of the other wave vector \( k_y = 0 \to 2\pi \) [17, 52]. While the Berry phase can only be defined for pure states, which precludes a direct application of this prescription to an open system, one can nevertheless define one or more geometric phases for a cyclic path of density matrices. The idea then is to use the winding numbers of these phases, which are quantized, as a topological characterization for open systems.

We will show that such a procedure can indeed be carried through. The
treatment hinges upon Uhlmann’s parallel transport of open systems [66, 35], to be summarized later, which maps a cyclic path of density matrices to a single matrix $M$; see Eq. 6.31 below. Geometric information about the path is to be extracted from $M$. The simplest prescription is Uhlmann’s phase, $\gamma \equiv \text{Arg}[\text{Tr} M]$, which has been extensively studied and is experimentally measurable [20, 1, 55, 81]. Recently Viyuela et al. [67] used $\tilde{\gamma}$ to identify topological transitions in 1D fermion systems at finite temperature, where $\tilde{\gamma}$ changes discretely from $\pi$ to 0 at a critical temperature $T_c$. For 2D systems, which are the focus of this chapter, we will compute $M$ at each $k_x$, and study the spectral flow of its (complex) eigenvalues with respect to $k_x$. This will be demonstrated in detail using Haldane’s honeycomb lattice model, where analytical treatment is possible. We will also present numerical results for the Bernevig-Hughes-Zhang [5] model as an example of the $Z_2$ topological insulators, and for the Hofstadter [29] model as an example of multi-band Chern insulators.

6.1 Two perspectives on Berry phase

Historically, Berry proposed geometric phase for pure quantum states under adiabatic evolution in terms of the cumulation of what is now known as the Berry connection [6]. While Berry’s formulation is more familiar to the condensed matter community, there is an alternative formulation due to Simon [60] using the language of parallel transport of wavefunctions, and it is along this line of thought that Uhlmann constructed a parallel transport for density matrices [66]. It is therefore instructive to first understand the connection between Berry’s and Simon’s perspectives for pure states.

Consider a two band Hamiltonian,

$$H_k = \cos \phi_k \sigma_x + \sin \phi_k \sigma_y . \tag{6.1}$$

Periodicity of $H_k$ in $k$ requires that $\phi_k$ winds integer times,

$$\phi(2\pi) - \phi(0) = 2\pi \cdot n , \quad n \in \mathbb{N} . \tag{6.2}$$

For concreteness we set $n = 1$. 

There are two ways to compute the Berry phase from wavefunctions. In the conventional way, one writes the energy eigenstates as

\[ |\psi^\pm_k\rangle = \frac{1}{\sqrt{2}} \left( \frac{1}{\pm e^{i\phi_k}} \right). \]  

(6.3)

Then the Berry connection and Berry phase of the + band are

\[ A_k = i \langle \psi_k^+ | \partial_k | \psi_k^+ \rangle = -\frac{1}{2} \partial_k \phi_k , \]  

(6.4)

\[ \gamma = \int_0^{2\pi} dk A_k = \frac{\phi(0) - \phi(2\pi)}{2} = -n\pi . \]  

(6.5)

The Berry phase is \( \pi \) times half the winding number of \( \phi(k) \). In this calculation, it is important to choose a phase convention for \( |\psi_k\rangle \) to be smooth in \( k \). A counterexample to this phase convention:

\[ |\tilde{\psi}^\pm_k\rangle = \frac{1}{\sqrt{2}} \left( e^{-i\phi_k/2} \pm e^{i\phi_k/2} \right), \]  

(6.6)

which is antiperiodic across \( k = 2\pi \).

An alternative way to extract the Berry phase is through the notion of parallel transport. Two states are defined to be parallel if their inner product is real [47]. Keeping one state fixed, the pair can always be made parallel by choosing a proper gauge for the second state. Starting with \( k = 0 \), say. We can fix the gauge at \( k = \delta, 2\delta, \cdots \), one by one, such that all neighboring pairs are parallel. A state at some \( k \) with its gauge fixed in this way is said to be parallel-transported from \( k = 0 \). Since to leading order in \( \delta \), the imaginary part of the overlap is the Berry connection itself, it must vanish along a parallel transport,

\[ \langle \tilde{\psi}_k^+ | i \partial_k | \tilde{\psi}_k^+ \rangle \overset{\text{parallel transport}}{=} 0 . \]  

(6.7)

This indeed is satisfied by Eq. 6.6. After a cyclic parallel transport, in this case from \( k = 0 \) to \( 2\pi \), the disagreement in the initial and final phases is the Berry phase. Thus the antiperiodicity of Eq. 6.6 in fact accounts for the Berry phase, \( -1 = e^{i\gamma} \).
6.2 Uhlmann’s parallel transport

In this section, we provide a self-contained review of Uhlmann’s parallel transport following Refs. [66, 35], and derive some useful results for later applications.

Consider an ordered set of density matrices $\rho_1, \rho_2, \cdots$ with $\rho_a = \rho(g_a)$ and $g_a$ is some parameter vector, e.g., a Bloch momentum. We do not impose the normalization $\text{Tr}(\rho_a) = 1$. Each $\rho_a$ can be lifted to a vector $W_a$ in a Hilbert-Schmidt space (i.e. the space of Hilbert-Schmidt operators),

$$W_a = \sqrt{\rho_a} U_a, \quad \rho_a = W_a W_a^\dagger,$$  

with arbitrary unitary $U_a$. The inner-product and norm in the Hilbert-Schmidt space is defined as

$$\langle W_a, W_b \rangle = \text{Tr}(W_a^\dagger W_b) \quad ||w|| = +\sqrt{\langle w, w \rangle}.$$  

6.2.1 Parallelity

Two vectors $W_a$ and $W_b$ are defined as parallel if the choice of $U_a$ and $U_b$ minimizes $||W_a - W_b||^2$, or equivalently, $\text{Tr}(W_a^\dagger W_b + h.c.)$ is maximal. This is attained if $W_a^\dagger W_b$ is hermitian. To see this, note that

$$\text{Tr}(W_a^\dagger W_b + h.c.) = \text{Tr}(\sqrt{\rho_a} \sqrt{\rho_b} U_a^\dagger U_b^\dagger + h.c.) = \text{Tr}(HV + V^\dagger H) = \text{Tr}[H(V + V^\dagger)]$$  

where $H$ and $V$ are defined through the polar decomposition

$$\sqrt{\rho_a} \sqrt{\rho_b} U_a U_b^\dagger = HV, \quad U_b U_a^\dagger = U_{ba} V, \quad H = H^\dagger \equiv \sqrt{\rho_a} \sqrt{\rho_b} U_{ba},$$  

with $U_{ba}$ and $V$ both unitary. Using the eigenbasis of $V = \sum_n \exp(i\theta_n) |v_n\rangle \langle v_n|$, one has $\text{Tr}[H(V + V^\dagger)] = 2 \sum_n \cos \theta_n \cdot \langle v_n|H|v_n\rangle$. $H$ is positive-semidefinite ($H = H^\dagger$) in band insulators a more appropriate object to use is the single particle correlation matrix $G$ which is not normalized, $G = f(\frac{H - \mu}{k_B T})$ where $f(x) = [\exp(x) + 1]^{-1}$ is the Fermi function and $H$ is the one-body Hamiltonian matrix. It plays the role of one-body “density matrix” in the sense that the expectation value of a one-body operator $O$ is $\langle O \rangle = \text{Tr}(GO)/\text{Tr}(G)$.\footnote{In band insulators a more appropriate object to use is the single particle correlation matrix $G$ which is not normalized, $G = f(\frac{H - \mu}{k_B T})$ where $f(x) = [\exp(x) + 1]^{-1}$ is the Fermi function and $H$ is the one-body Hamiltonian matrix. It plays the role of one-body “density matrix” in the sense that the expectation value of a one-body operator $O$ is $\langle O \rangle = \text{Tr}(GO)/\text{Tr}(G)$.}

\footnote{The same requisite on a pair of pure states enforces their inner product to be real. For two infinitesimally different pure states, their Berry connection, being the imaginary part of leading order, must thus vanish.} The same requisite on a pair of pure states enforces their inner product to be real. For two infinitesimally different pure states, their Berry connection, being the imaginary part of leading order, must thus vanish.
\[ \sqrt{\rho_a} \sqrt{\rho_b} \], cf. Eq. 6.13 below), thus \( \langle v_n | H | v_n \rangle \geq 0 \), so to maximize the trace, \( \theta_n = 0 \), hence \( V = I \) and \( W_a^\dagger W_b = U_a^\dagger H U_a \) is hermitian.\(^3\)

### 6.2.2 Properties of parallel vectors

We collect below some notable properties for a parallel pair \( W_a \) and \( W_b \). Uhlmann’s relative phase factor between \( W_a \) and \( W_b \), \( RPF(\rho_a,\rho_b) \), is

\[ U_{ab} \equiv U_a U_b^\dagger = (U_{ba})^\dagger. \tag{6.12} \]

From Eq. 6.11 and \( H = \sqrt{HH^\dagger} \), one has the polar decomposition

\[ \sqrt{\rho_a} \sqrt{\rho_b} = ||\sqrt{\rho_a} \sqrt{\rho_b}|| \cdot U_{ab} = U_{ab} \cdot ||\sqrt{\rho_a} \sqrt{\rho_b}||, \tag{6.13} \]

where the second equality follows from the hermiticity of \( U_{ab} (W_a^\dagger W_b) U_b^\dagger \), and by definition

\[ F(\rho_a,\rho_b) \equiv ||\sqrt{\rho_a} \sqrt{\rho_b}|| = \sqrt{\rho_a \rho_b \rho_a}, \tag{6.14} \]

\( F \) is known as the fidelity (which Uhlmann termed \( AMP(\rho_a,\rho_b) \)). Note that \( \sqrt{\rho_a} \sqrt{\rho_b} \neq \sqrt{\rho_a \rho_b} \) since \( \rho_a \) and \( \rho_b \) don’t necessarily commute. In fact, using the hermiticity \( W_b^\dagger W_a = W_a^\dagger W_b \), one has \( (W_a W_b^\dagger)^2 = W_a W_a^\dagger W_b W_b^\dagger = \rho_a \rho_b \), viz.,

\[ W_a W_b^\dagger = \sqrt{\rho_a \rho_b} = \sqrt{\rho_a} U_{ab} \sqrt{\rho_b}, \tag{6.15} \]

where the second equality follows from the definition Eq. 6.8\(^4\).

The connection \( U_{ab} \) is invariant under rescaling of the density matrices \( \rho_a \to \lambda_a \rho_a \) where \( \lambda_a \in \mathbb{R} \) can be different for different \( a \). In particular, imposing a normalization \( \text{Tr} \rho_a = 1 \) does not change \( U_{ab} \).

Eqs. 6.13 and 6.15 are general and hold even for \( \rho_{a,b} \) with zero eigenvalues — unfaithful in Uhlmann’s language. If \( \rho_{a,b} \) are faithful, then \( U_{ab} \) can be explicitly

\(^3\)In Ref. [20], above eq (3), the parallel transport is defined as “each \(|\langle w_t, w_{t'} \rangle|\) is constrained to its maximum ... \( \text{Tr} \sqrt{\rho_a} \rho_{t'} \sqrt{\rho_a} \).” This translates to maximizing \( \text{Tr}(W_a^\dagger W_b) \times \text{c.c.} \) instead of Uhlmann’s original definition where \( \text{Tr}(W_a^\dagger W_b) + \text{c.c.} \) is maximized. The two are slightly different: Still using the same polar decomposition as in the text, \( W_a^\dagger W_b = HV \), then \( \text{Tr}(HV) \times \text{c.c.} = \sum_{m,n} H_m H_n e^{i(\theta_m - \theta_n)} = \sum_m H_m^2 + 2 \sum_{m<n} H_m H_n \cos(\theta_m - \theta_n) \) where \( H_m \equiv \langle v_m | H | v_m \rangle \) is non-negative. Then maximization w.r.t \( \theta \) yields all \( \theta_m \) equal, viz., \( V = e^{-i\theta} I \).

\(^4\)Note that \( \rho_a \rho_b \) could have complex eigenvalues, which may require care when taking \( \sqrt{\cdots} \).
written as

\[
U_{ab} = ||\sqrt{\rho_a} \sqrt{\rho_b}| |^{-1} \cdot \sqrt{\rho_a} \sqrt{\rho_b} = \sqrt{\rho_a} \sqrt{\rho_b} \cdot ||\sqrt{\rho_b} \sqrt{\rho_a}| |^{-1}
\]

\[
= ||\sqrt{\rho_a} \sqrt{\rho_b}| | \cdot \sqrt{\rho_a^{-1}} \sqrt{\rho_b^{-1}} = \sqrt{\rho_a^{-1}} \sqrt{\rho_b^{-1}} ||\sqrt{\rho_b} \sqrt{\rho_a}| |
\]

(6.16)

where the first line follows from Eq. 6.13 and the second line from \( U_{ab} = (U_{ab}^{-1})^\dagger \)

### 6.2.3 Constructing \( U_{ab} \) by singular value decomposition

Let \( \rho_a \) be a \( N \times N \) matrix with \( \nu_a \) nonzero eigenvalues, for example, \( \nu_a \) filled bands at Bloch momentum \( k_a \), with the total number of bands fixed at \( N \). Assume \( \sqrt{\rho_a} \sqrt{\rho_b} \) has the following SVD,

\[
\sqrt{\rho_a} \sqrt{\rho_b} = L_a D_{ab} R_b^\dagger \quad \text{(no sum)} \quad (6.17)
\]

where \( L_a \) and \( R_b \) are oblong \( N \times \nu_{a,b} \) unitary matrices and \( D_{ab} \) is a \( \nu_{ab} \times \nu_{ab} \) diagonal matrix with non-negative entries, \( \nu_{ab} \) should be \( \min(\nu_a, \nu_b) \). The meaning of unitarity is

\[
L_a L_a^\dagger = R_a R_a^\dagger = P_a \quad \text{(Projection onto support of } \rho_a \text{)} , \quad (6.18)
\]

\[
L_a^\dagger L_a = R_a^\dagger R_a = 1_{\nu_a \times \nu_a} \quad \text{(orthonormality)} . \quad (6.19)
\]

The SVD is unique up to a \( \nu_{ab} \times \nu_{ab} \) gauge transformation \( \Lambda \) under which \( D_{ab} \) is invariant,

\[
L_a \rightarrow L_a \Lambda , \quad R_b \rightarrow R_b \Lambda , \quad \Lambda D_{ab} \Lambda^\dagger = D_{ab} . \quad (6.20)
\]

For non-degenerate \( D_{ab} \), \( \Lambda \) is a diagonal matrix of phase factors. Then Eq. 6.14 becomes

\[
||\sqrt{\rho_a} \sqrt{\rho_b}| | = \sqrt{L_a (D_{ab})^2 L_a^\dagger} = L_a D_{ab} L_a^\dagger \quad \text{(no sum)} . \quad (6.21)
\]

Substituting this into Eq. 6.13 then yields

\[
U_{ab} = L_a R_b^\dagger . \quad (6.22)
\]

\footnote{Note that the indices \( a, b \) on \( L \) and \( R \) are only for notational convenience and do not mean, for example, that \( L_a \) depends on \( \rho_a \) only. In fact \( L_a \) diagonalizes \( \sqrt{\rho_a} \rho_b \sqrt{\rho_a} \). \( R_b \) diagonalizes \( \sqrt{\rho_b} \rho_a \sqrt{\rho_b} \), both depending on \( \rho_a \) and \( \rho_b \).}
Note that this result is invariant under Λ. $U_{ab}$ is of size $N \times N$.

In the above derivation one may worry about the signs of $\sqrt{(D_{ab})^2}$. But note that one may as well directly rewrite $\sqrt{\rho_a} \sqrt{\rho_b} = L_a D_{ab} L_a^\dagger \times L_b R_b^\dagger$. Since $L_a D_{ab} L_a^\dagger$ is positive-semidefinite hermitian and $L_b R_b^\dagger$ is unitary, this is a polar decomposition, and by uniqueness the decomposition.

**Ambiguity from zero singular values**

So far we have excluded zero singular values by allowing oblong $L$ and $R$ matrices. In practice, zeros may come as tiny numbers. Let’s use the other convention of SVD where $L_a$, $D_{ab}$ and $R_b$ are all $N \times N$ matrices, with a subset of $D_{ab}$ identically “zero”. Then instead of Eq. 6.20, one has

$$L_a \to L_a \left( \Lambda_{KL} \right), \quad R_b \to R_b \left( \Lambda_{KR} \right),$$

$$\left( \Lambda^\dagger_{KL} \right) \left( \tilde{D}_{ab} \right) \left( \Lambda_{KR} \right) = \left( \tilde{D}_{ab} \right) \equiv D_{ab} \quad (6.24)$$

where $\Gamma$ is a unitary transformation within the support of $||\sqrt{\rho_a} \sqrt{\rho_b}||$, and $K_{L,R}$ are two unrelated unitary transformation within the kernel of $||\sqrt{\rho_a} \sqrt{\rho_b}||$. The connection $U_{ab}$ under the transformation becomes

$$L_a R_b^\dagger \to L_a \left( \mathbb{I} \atop U_K \right) R_b^\dagger, \quad U_K \equiv K_{L}^\dagger K_{R},$$

i.e., there is a $U_K$ ambiguity. As we shall see, this ambiguity will go away when projected by $\rho$ (which, at zero temperature, is a projector), see discussion beneath Eq. 6.49.

**6.2.4 Purification**

A purification maps the amplitude $W_a$ of the density matrix $\rho_a$ to a pure state $|W_a\rangle$ in a larger Hilbert space. Writing

$$W_a = \sum_n \sqrt{p_a^n} |\psi_a^n\rangle \langle \phi_a^n|, \quad |\phi_a^n\rangle = U_a^n |\psi_a^n\rangle \quad (6.26)$$

one then constructs (note the complex conjugation on $|\phi\rangle$)

$$|W_a\rangle = \sum_n \sqrt{p_a^n} |\psi_a^n\rangle \otimes |\phi_a^n|^\star \quad (6.27)$$
such that the reduced density matrix of $|W_a\rangle$ is $\rho_a$, and the inner product is preserved,

$$(W_a, W_b) \equiv \text{Tr}(W_a^\dagger W_b) = \frac{1}{2} \langle W_a|W_b \rangle = \sum_{mn} \sqrt{p_a^m p_b^n} \langle \psi^m_a|\psi^n_b \rangle \langle \psi^m_a|U_{ab}|\psi^n_b \rangle^* .$$  \hspace{1cm} (6.28)

### 6.2.5 Parallel transport and Holonomy

A parallel lift $\{W_0, W_1, W_2, \cdots, W_N\}$ of the density matrices $\{\rho_0, \rho_1, \rho_2, \cdots, \rho_N\}$ is one where all neighboring $W_a$’s are parallel. This induces a parallel transport from $U_0$ to $U_N$,

$$U_{0N} = U_{01}U_{12}\cdots U_{N-1N} \hspace{1cm} (6.29)$$

$$= L_0 R_1^\dagger L_1 R_2^\dagger \cdots L_{N-1} R_N^\dagger , \hspace{1cm} (6.30)$$

which depends only on the $\rho$’s following Eq. 6.16.

The physically interesting case is when one has a loop, $\rho_0 = \rho_N$. $W_0$ and $W_N$ are in general different and not parallel. Their disagreement can be characterized by a holonomy matrix defined as

$$M_0 \equiv \rho_0 U_{0N} . \hspace{1cm} (6.31)$$

In the pure case, this reduces to $\langle \psi_0|\psi_N \rangle$ which gives the Berry phase. Note that

$$W_N^\dagger W_0 = U_N^\dagger \rho_0 U_0 \implies \text{Spec}(W_N^\dagger W_0) = \text{Spec}(\rho_0 U_{0N}) \hspace{1cm} (6.32)$$

where to get the second equality we used $\text{Spec}(AB) = \text{Spec}(BA)$, i.e. the cyclic invariance of the spectrum.

The args of $\text{Spec}(\rho_0 U_{0N})$ must come in $\pm$ pairs. This is because $\text{Det}(L_0 R_1^\dagger) = \text{Det}(\sqrt{\rho_0} \sqrt{\rho_1})/\text{Det}(D_{01}) \in \mathbb{R}$ — as a consequence, $\text{Det}(\rho_0 U_{0N})$ is real.

In terms of purification,

$$(W_N, W_0) = \langle W_N|W_0 \rangle = \text{Tr}(M_0) \equiv v_0 e^{i\gamma_0} . \hspace{1cm} (6.33)$$

Here $\gamma_0$ is a geometric phase\(^6\) and $v_0$ is its visibility. The “$0$” subscript denotes their dependence on the point of observation in the loop ($a = 0$ in this case). One

\(^6\)This is the usual definition of the Uhlmann phase in the literature although Uhlmann himself never seems to have defined such an object.
might take the product \((W_{a+N}, W_a)\) for any \(a\), assuming parallelity is enforced along \(W_a, W_{a+1}, \ldots, W_{N-1}, W_{N+0}, W_{N+1}, \ldots W_{N+a}\), then in principle \(v_a\) and \(\gamma_a\) can be different for different \(a\). One also loses the matrix structure of \(U_{0N}\) in the purification picture.

### 6.2.6 Connection with band projectors

We show below that at \(T = 0\) Uhlmann’s phase reduces to Wannier center drifts in band insulators.

**Single occupied band**

Assume \(\rho_a = |\psi_a\rangle \langle \psi_a| \equiv P_a\), then using \(P_a = P_a^2 = \sqrt{P_a}\), Eq. 6.13 becomes

\[
P_a P_b = \sqrt{P_a P_b P_a} U_{ab},
\]

where \(\sqrt{P_a P_b P_a} = \sqrt{|\psi_a\rangle \langle \psi_b| \langle \psi_a| \langle \psi_a|} = |\psi_a\rangle \langle \psi_b|,\)

thus\(^7\)

\[
U_{ab} = \frac{P_a P_b}{|\langle \psi_a| \psi_b\rangle|}, \tag{6.35}
\]

assuming neighboring states are not orthogonal. This is exactly the Berry connection. Thus the single nonzero eigenvalue of \(U_{0N}\) is \(e^{i\gamma}\) where \(\gamma\) is the Berry phase.

**Multiple equal-weight bands**

Consider \(\nu\) occupied bands of a \(N\)-band system with equal weights,

\[
\rho_k = \frac{1}{\nu} \sum_{n=1}^{\nu} |\psi_{n,k}\rangle \langle \psi_{n,k}| = \frac{1}{\nu} G_k \tag{6.36}
\]

\(^7\)The exact relation is

\[
P_a U_{ab} = \frac{P_a P_b}{|\langle \psi_a| \psi_b\rangle|} \tag{6.34}
\]

According to Uhlmann, for unfaithful \(\rho\)’s, \(U_{ab}\) is such that its left support is the same as that of \(\rho_a\) and its right support that of \(\rho_b\), thus \(P_a U_{ab} = U_{ab} = U_{ab} P_b\). This corresponds to picking a convention in SVD such that all singular values are nonzero (other than accidental zeros) and the two unitaries are oblong.
where $G_k$ is the projector onto the $\nu$ bands. The SVD Eq. 6.17 becomes

$$\sqrt{p_k}\sqrt{p_{k+dk}} = \frac{1}{\nu} G_k G_{k+dk} = L_k D_{k,k+dk} R_{k+dk}^\dagger . \quad (6.37)$$

For infinitesimal $dk$, $G_k G_{k+dk}$ is unitary, \(^8\) hence its nonzero singular values are all 1, implying

$$D_{k,k+dk} = \frac{1}{\nu} I_{\nu \times \nu} + O(dk^2) , \quad U_{k,k+dk} = G_k G_{k+dk} . \quad (6.40)$$

Thus the non-abelian Uhlmann phases for equal-weight bands are identical to the non-abelian Wannier drifts.

It is instructive to formulate it in the other SVD convention, i.e. when all matrices are $N \times N$ and singular values could be zero. Write

$$\xi_k = (|\psi^1_k\rangle, |\psi^2_k\rangle \cdots |\psi^\nu_k\rangle) \quad (6.41)$$

where $|\psi^a_k\rangle$ is the $a^{th}$ occupied state, an $N$-element vector. Then

$$G_k = \xi_k \xi_k^\dagger . \quad (6.42)$$

A numerical SVD of $G_a G_b$ would yield

$$G_a G_b = (\xi_a, 0) \left( \begin{array}{c} \xi^\dagger_a \\ 0 \end{array} \right) (\xi_b, 0) \left( \begin{array}{c} \xi^\dagger_b \\ 0 \end{array} \right) = (\xi_a, 0) \left( \begin{array}{cc} \xi^\dagger_a \xi_b \\ 0 \end{array} \right) \left( \begin{array}{c} \xi^\dagger_b \\ 0 \end{array} \right) \quad (6.43)$$

$$= (\xi_a, \tilde{\xi}_a) \left( \begin{array}{c} I \\ 0 \end{array} \right) \left( \begin{array}{c} \xi^\dagger_a G_b \\ \tilde{\xi}^\dagger_b \end{array} \right) \quad (6.44)$$

where $\tilde{\xi}$'s are $N \times (N - \nu)$ matrices the columns of which form bases in the unoccupied states,

$$G_a \tilde{\xi}_a = G_b \tilde{\xi}_b = 0 . \quad (6.45)$$

\(^8\)Writing

$$\langle \psi_m, k | \psi_n(k + dk) \rangle = \left( e^{iA(k)\cdot dk} \right)_{mn} + O(dk^2) , \quad A_{mn}(k) = -i \langle \psi_m, k | \partial_k \psi_n, k \rangle \quad (6.38)$$

where $A(k)$ is the non-abelian Berry connection, one has that

$$G_k G_{k+dk} = \sum_{m,n=1}^\nu \left( e^{iA(k)\cdot dk} \right)_{mn} \langle \psi_m, k | \psi_n, k+dk \rangle + O(dk^2) \quad (6.39)$$

which is manifestly unitary.
One has

\[ U_{ab} = G_a G_b + \tilde{\xi}_a \tilde{\xi}_b^\dagger, \quad (6.46) \]

\[ U_{ab} U_{bc} = (G_a G_b + \tilde{\xi}_a \tilde{\xi}_b^\dagger)(G_b G_c + \tilde{\xi}_b \tilde{\xi}_c^\dagger) = G_a G_b G_c + \tilde{\xi}_a \tilde{\xi}_c^\dagger, \quad (6.47) \]

\[ \ldots \]

\[ U_{0N} = G_0 G_1 \cdots G_N + \tilde{\xi}_0 \tilde{\xi}_N^\dagger. \quad (6.49) \]

Upon completing the cyclic loop and weighting by the density matrix at the observation point, the transport in the unoccupied space \( \tilde{\xi}_0 \tilde{\xi}_N^\dagger \) is projected out, leaving only the Wilson loop.

### 6.2.7 2-band models

A \( 2 \times 2 \) matrix \( M \) satisfies \( M^2 - M \text{Tr}M + \text{Det}M = 0 \).\(^9\) Thus

\[ M = \frac{M^2 + \text{Det}M}{\text{Tr}M}, \quad \text{Tr}M = \sqrt{\text{Tr}(M^2 + \text{Det}M)} . \quad (6.50) \]

The second equation is obtained from taking the trace of the first equation. Now let \( M = \sqrt{\rho_1 \rho_2} \). \( M \) is positive-semidefinite. Further denote

\[ \sqrt{\rho_i} = a_i + b_i \cdot \sigma. \quad (6.51) \]

Using \( \sqrt{\rho_i}^{-1} = (a_i - b_i \cdot \sigma) / \text{Det} \sqrt{\rho_i} \), Eq. 6.16 yields\(^10\)

\[ U_{12} = \frac{\sqrt{\rho_1} \sqrt{\rho_2} + \text{Det} \sqrt{\rho_1} \text{Det} \sqrt{\rho_2} \rho_1^{-1} \sqrt{\rho_2}^{-1}}{\sqrt{\text{Tr} (\rho_1 \rho_2 + \text{Det} \rho_1 \text{Det} \rho_2)}} = \frac{a_1 a_2 + b_1 \cdot b_2 + i(b_1 \times b_2) \cdot \sigma}{\sqrt{(a_1 a_2 + b_1 \cdot b_2)^2 + (b_1 \times b_2)^2}} . \quad (6.52) \]

The RHS is explicitly unitary. One can verify that

\[ U_{k,k+\delta k} = 1 + i \delta k \frac{b_k \times \nabla_k b_k}{|b_k|^2} \cdot \sigma + O(\delta k^2) . \quad (6.53) \]

\(^9\)In the diagonal basis this is simply the characteristic equation for eigenvalues, \( x^2 - x \text{Tr}M + \text{Det}M = 0 \).

\(^10\)\( U_{ab} \) is ambiguous in the sense of §6.2.3 if \( ||\sqrt{\rho_a} \sqrt{\rho_b}|| \) has zero eigenvalues — e.g., if both \( \rho_i \) are projectors. In the present case, \( U_K \) of Eq. 6.25 is a unimodular number and the solution Eq. 6.52 amounts to choosing a convention for \( U_K \) such that \( \text{Det}(U_{ab}) \in \mathbb{R} \).
Enforcing a normalization of $\rho_k$ amounts to simultaneously rescaling $a_k$ and $b_k$, under which the result is invariant. Upon completing a Wilson loop, one obtains

$$U_{k,k+2\pi} = \mathcal{P} \exp \left[ -i \oint_k d\tilde{b}_k \times \tilde{b}_k \cdot \sigma \right], \quad \tilde{b}_k \equiv \frac{b_k}{\sqrt{a_k^2 + b_k^2}}$$

(6.54)

where $\mathcal{P}$ denotes path ordering.

**Projector limit**

In the limit $a_k = |b_k| = \frac{1}{2}$, e.g. at zero temperature, $\{\rho_k\}$ are projectors, and one has

$$U_{k,k+2\pi} = \mathcal{P} \exp \left[ -\frac{i}{2} \oint d\hat{b}_k \times \hat{b}_k \cdot \sigma \right],$$

(6.55)

where $\hat{b}_k = b_k/|b_k|$.

**Coplanar limit**

If the vectors $b_k$ for different $k$ are coplanar, e.g., in the SSH model or at the edge crossing point in the Haldane model, then one can drop the path ordering and get

$$U_{k,k+2\pi} = \exp (i2S_{\hat{b}} \sigma_{\perp}) \quad , \quad S_{\hat{b}} = \frac{1}{2} \hat{n}_{\perp} \cdot \oint \tilde{b}_k \times d\tilde{b}_k$$

(6.56)

where $S_{\hat{b}}$ is the directed area spanned by the path of $\tilde{b}_k$ and $\sigma_{\perp}$ is the Pauli matrix along the normal direction of the plane. Note that $S_{\hat{b}}$ is independent of the observation point $k$. In the projector limit, $\tilde{b}_k$ lies on a circle of radius $\frac{1}{\sqrt{2}}$ (cf. Eq. 6.54), thus $U_{k,k+2\pi} = \exp (in\pi)$ where $n$ is the winding number of $\hat{b}_k$.

The spectrum of the holonomy matrix $M_k = \rho_k U_{k,k+2\pi}$ can be solved: rotating to the frame where $b_k$ is in the $xy$ plane, then

$$\rho_k = \frac{1}{2} \text{Tr}(\rho_k) \begin{pmatrix} 1 & z_k^* \\ z_k & 1 \end{pmatrix}, \quad U_{k,k+2\pi} = \begin{pmatrix} e^{2iS} & 0 \\ 0 & e^{-2iS} \end{pmatrix},$$

(6.57)

where $z_k$ is some complex number with $|z_k| = 1$ at $T = 0$ and $z_k = 0$ at $T \to \infty$. More explicitly

$$|z| = \frac{f_> - f_<}{f_> + f_<},$$

(6.58)
where \( f \) are the two Fermi occupation numbers in \( \rho \). The spectrum of \( M_k \) is thus
\[
m_k^\pm = \frac{1}{2} \text{Tr}(\rho_k) \left[ \cos(2S) \pm \sqrt{|z|^2 - \sin^2(2S)} \right].
\] (6.59)

For \( |z| < \sin(2S) \), \( m_k^\pm \) form a complex conjugate pair.

Taking the sum of the eigenvalues yields the inner product trace, \( \text{Tr}(\rho_k U_{k,k+2\pi}) \), which is independent of the observation point \( k \) if \( \text{Tr}(\rho_k) \) is normalized to a constant,
\[
\text{Tr}(\rho_k U_{k,k+2\pi}) = \text{Tr}(\rho_k) \cos \left[ 2S_b \right].
\] (6.60)

As temperature increases, the area \( S_b \) shrinks from \( \pi/2 \) to zero, and the inner product interpolates between \(-1\) and \(1\). According to the definition Eq. 6.33, the Uhlmann (trace) phase \( \gamma \) changes abruptly from \( \pi \) to 0 when \( S_b \) shrinks to \( \pi/4 \), i.e. half its zero temperature size.

### 6.3 Topological characterization of 2D insulators at finite temperature

We formulate our scheme for general 2D \( N \)-band insulators with \( L_x \times L_y \) unit cells, assuming translational invariance and periodic boundary conditions in both directions. Let \( |\psi_{nk}\rangle \) be the eigenstates of the momentum space Hamiltonian \( H_k \), where \( n \) is the band index and \( \mathbf{k} \) is the Bloch momentum. At the single particle level, the role of density matrix is played by the correlation matrix,
\[
\rho_k(\mu, T) = \sum_{n=1}^{N} f_{nk}(\mu, T) |\psi_{nk}\rangle \langle \psi_{nk}|
\] (6.61)

where \( f_{nk}(\mu, T) \) is the Fermi distribution and \( \mu \) is the chemical potential. Note that \( \rho_k \) is a \( N \times N \) matrix, and it is not normalized, \( \text{Tr} \rho_k \neq 1 \). At fixed \( k_x \), one can compute the holonomy \( M_{k_x}(\mu, T) \) over the path \( k_y \in [0, 2\pi] \). Then as \( k_x \) sweeps a \( 2\pi \) cycle, the eigenvalues \( \{m_i\} \) trace out closed paths in the complex plane.

The topological numbers of the system are to be defined from the winding numbers of the phases \( \{\gamma_i(k_x)\} \). However, we need to take into account that
the amplitude spectrum \( \{ r_i(k_x) \} \) has a gap structure much like the energy band spectrum. If a particular level \( r_i(k_x) \) is gapped from the rest, then one can define a winding number of the corresponding geometric phase, \( C_i \equiv [\gamma_i(2\pi) - \gamma_i(0)]/2\pi \). At zero temperature, since \( M \) reduces to the Wilson loop operator, \( C_i \) reduces to the winding number of the \( i^{th} \) non-Abelian Berry phase [32]. If several levels – e.g. \( r_1, r_2 \) and \( r_3 \) – evolve into each other but are otherwise separated from the rest, then their winding has to be considered in a collective manner. For Chern insulators, two natural choices arise: (1) The collective topological number could be the winding number of the sum of the phases, \( \gamma_1 + \gamma_2 + \gamma_3 \). This choice is motivated by the analogy with zero temperature gapless energy bands, where the total Chern number is the winding number of the sum of the individual Berry phases. (2) Alternatively it could also be defined as the winding number of the phase of the sum, \( \text{Arg}(m_1 + m_2 + m_3) \). Such a choice draws analogy from multi-path interference type experiments, where each complex eigenvalue \( m_i \) encodes both the weight and the phase of the \( i^{th} \) path, and the output is a coherent sum of these complex weights. In both cases, with \( g \) spectral gaps in \( \{ r_i \} \), one obtains \( g + 1 \) topological numbers. For time reversal invariant topological insulators, one should instead consider the time-reversal partner switching similar to that of the non-Abelian Berry phases [22, 80]. As temperature increases from zero, the gap structure of \( \{ r_i \} \) also changes, and the system experiences a series of topological transitions until it reaches a fully trivial stage where all topological numbers are zero.

We should point out that Eq. 6.31 is defined in terms of a cyclic parallel transport initiating at particular point \( g_0 \). One may associate, to each \( g \) on the cyclic path, a holonomy matrix \( M_a = \rho_a U_a U_{a+N}^\dagger \), whose spectra may differ. The general picture of temperature-driven amplitude gap evolution and hence topological transitions remains unchanged, although the particular temperatures at which the gap structure changes will vary with \( a \). In certain cases, such as the Haldane model to be discussed later, an \( a \)-independent transition temperature could nonetheless be defined.

\(^{11}\)For case (1), only \( g \) of them are independent from the \( \text{Det}(M) \) restriction
Figure 6.1: Uhlmann spectral flow of the Haldane model. Top panels: amplitudes of eigenvalues of $M$. Bottom panels: phases of eigenvalues, black solid lines correspond to the Berry phase of the lower band (i.e., $T = 0$ limit). Color/point type encodes the eigenvalue index. In (a), the amplitudes are gapped, and the two phases wind in opposite directions. In (b) and (c), the amplitudes are gapless. The phases do not wind. In (d), the amplitudes are gapped again. The phases do not wind. Parameters used: $m = 0.5, \phi = 0.3\pi, t = 0.3, L_x = 200, L_y = 50, \mu = 0.5$.

6.4 Haldane model

The Haldane model [25] describes electrons hopping on a honeycomb lattice in a fluctuating magnetic field. The model has three parameters $t, m, \phi$, where $t \equiv t_{\text{NNN}}/t_{\text{NN}}$ is the ratio of hopping amplitudes between second neighbors and first neighbors, $m$ is the Semenoff mass contrasting the two sublattices, and $\phi$ is a phase associated with the second neighbor hops which breaks time reversal symmetry. The model can either be a Chern insulator (with $C = \pm 1$) or a trivial insulator ($C = 0$) depending on the choice of parameters.

In Fig. 6.1, we plot the spectral flow of holonomy eigenvalues (both amplitudes and phases) as functions of $k_x$, at three different temperatures. Here $k_x \equiv \mathbf{k} \cdot \mathbf{a}_x$ is the Bloch wave vector along the honeycomb basis vector in the $x$ direction. The matrix $M$ has two eigenvalues for two band models. From $\text{Det} M > 0$, the two geometric phases are opposite to each other, and we shall focus on $\gamma_>$, the
We find that there are three temperature regimes with distinctive spectral patterns: (i) In the low temperature regime (panel a), $\gamma_>$ winds once, and its spectral flow is a minor deviation from the Berry phase flow (i.e. its zero temperature limit). The two amplitudes remain gapped. As $T$ increases, the amplitude gap reduces and the deviation of $\gamma_>$ increases. (ii) In the intermediate temperature regime (panels b and c), the amplitudes touch and stay gapless. There is no winding in the individual phases. (iii) In the high temperature regime (panel d), the amplitudes are gapped again, and $\gamma_>$ does not wind.

Two-band models yield to analytical treatment [35] useful for understanding this temperature trichotomy. At fixed $k_x$, the correlation matrices are labeled by the Bloch wave vector along the $a_2$ basis vector of the honeycomb lattice, denoted as $k = k \cdot a_2$. Since these are $2 \times 2$ matrices, we can parameterize

$$\sqrt{\rho_k} = \chi_k \left[ \sqrt{1 - |b_k|^2} + b_k \cdot \sigma \right],$$

which implicitly defines $\chi_k$ and $b_k$ through Eq. 6.61. Here $\sigma = (\sigma_x, \sigma_y, \sigma_z)$ are the three Pauli matrices. In Fig. 6.1, the $k_x$ point at which the amplitude gap closes – denoted as $k_c$ – coincides with where the $\pi$ Berry phase occurs. This is in fact the coplanar point, $k_c = -\sin^{-1}(m/6t \sin \phi)$ [31], where the entire path of $b_k$ lies on the same plane that passes through the origin.

The winding of $\gamma_>$ over $k_x$ is entirely determined from its value at $k_x = k_c$: it winds once if $\gamma_>(k_c) = \pi$, otherwise it does not wind. The eigenvalues of $M(k_c)$ are

$$m^\pm = \frac{1}{2} \text{Tr}(\rho_0) \left[ \cos(2S) \pm \sqrt{z^2 - \sin^2(2S)} \right],$$

with $\rho_0$ the correlation matrix at $(k_x, k_y) = (k_c, 0)$, and

$$S = \frac{1}{2} \left| \oint_k b_k \times db_k \right|, \quad z = \frac{f_> - f_<}{f_> + f_<},$$

i.e. $S$ is the area enclosed by the path of $b_k$. Here $f_> \geq f_<$ are the two eigenvalues of $\rho_0$ (i.e. Fermi weights). From Eqs. 6.63 and 6.64, we can understand the spectral...
evolution as a function of temperature: If $z < |\sin(2S)|$, $m^\pm$ form a complex conjugate pair, and the amplitude spectrum is gapless; otherwise they are both real and the amplitudes are gapped. In the gapped regimes, one can check the zero and infinite temperature limits: at $T = 0$, $S = \pi/2$ and $z = 1$, yielding $m^+ = 0$ and $m^- = -1$, hence $\gamma_>(T = 0) = \pi$, whereas for $T \to +\infty$, $z = S = 0$, $m^+ = 1$ and $m^- = 0$, hence $\gamma_>(T \to \infty) = 0$.

As discussed before, in the regime with gapless amplitudes, one has to consider the winding of a collective phase. Using $\tilde{\gamma} \equiv \text{Arg}[\text{Tr} M]$ (approach 2 in previous discussion) and Eq. 6.63, we have, at $k_x = k_c$,

$$\exp(i\tilde{\gamma}) = \text{Sgn}[\cos(2S)]$$

(6.65)

On the other hand, in the gapped regimes, it follows from $\text{Det} M > 0$ that $m^\pm$ and hence $\text{Tr} M$ must have the same sign, implying that $\gamma_> = \tilde{\gamma}$ at $k_c$. Thus one can use $\tilde{\gamma}$ in the entire range of $T$ as the geometric phase. The topological index is entirely determined by $\tilde{\gamma}$ at $k_c$, or equivalently the area $S$ circulated by the path of $b_k$, regardless of the gap structure of the amplitudes – a feature of two band models. Since increasing $T$ generically causes the loop area $S$ to shrink, $\text{Sgn}[\cos(2S)]$ must change from “−” to “+” and not the other way around. One can thus define a unique transition temperature $T_c$ such that $\tilde{\gamma}(k_c)$ changes discretely from $\pi$ to 0. The 2D topological transition coincides with the effective 1D transition [67] at $k_c$. Geometrically, such a topological transition occurs when the loop area $S$ reaches half of its zero temperature value, $S(T_c) = S(0)/2 = \pi/4$. Note that the same critical temperature $T_c$ would be obtained by using $M_a = \rho_a U_a U_{a+N}^T$ for $a \neq 0$, because $S$ is independent of $a$ according to Eq. 6.64.

The transition temperature $T_c$ may depend on other external parameters of the system as well. In Fig. 6.2, we allow the chemical potential $\mu$ to vary, and plot $\text{Tr} M = \cos(2S)$ in the $(\mu, T)$ space. $T_c(\mu)$ is determined as the curve at which $\cos(2S)$ crosses zero.

---

12 At $T = 0$, $|b_k| = 1/\sqrt{2}$, which follows from the projective nature of $\rho_k$ (and hence $\sqrt{\rho_k}$) at $T = 0$. 
Figure 6.2: \(\text{Tr} M\) at \(k_x = k_c\) for the Haldane model in the parameter space of temperature \(T\) and chemical potential \(\mu\). At \(k_x = k_c\), \(\text{Tr} M\) is real so \(\tilde{\gamma}\) depends solely on its sign. A topological transition occurs when \(\text{Tr} M\) crosses zero, giving \(T_c\) as a function of \(\mu\). Other parameters are the same as Fig. 6.1.
Figure 6.3: Uhlmann spectral flow of the BHZ model. Top panels: amplitudes of eigenvalues of $M$. Bottom panels: phases of eigenvalues. Color/point type encodes the eigenvalue index. The amplitudes are gapped in (a) and (d), but gapless in (b) and (c). Partner switching occurs in (a) but not in (d), and is ambiguous for the gapless (b) and (c), see text. Parameters used: $m = 1.1$, $\Delta = 0.3$.

6.5 BHZ model

We study the BHZ model [5] as an example of $Z_2$ TIs. The Hamiltonian is

$$H = \sin k_x \sigma_z \tau_x + \sin k_y \tau_y + (2 - m - \cos k_x - \cos k_y)\tau_z + \Delta \sigma_y \tau_y$$

where $\tau$ and $\sigma$ are two sets of Pauli matrices, and $\Delta \neq 0$ breaks inversion symmetry. The parameters will be chosen so that it is in the topological phase at zero temperature. In Fig. 6.3, we plot its holonomy spectrum in half a Brillouin zone – the other half contains redundant information due to time reversal symmetry. The four amplitudes form two pairs (partners), which are gapless for intermediate temperature but gapped for both low and high temperatures. In the low temperature regime (Panel a), each pair of the geometric phases exhibit partner switching [22, 80] where they change from $\pi$ at $k_x = 0$ to 0 at $k_x = \pi$, hence the system is topological. After going through the gapless regime (b) and (c), the system becomes trivial in the high temperature gapped regime (d) where there is no partner switching.
6.6 Hofstadter model

The Haldane model is special in that it has only two bands, where the two geometric phases are not independent but must be opposite to each other as required by $\text{Det} M > 0$. The situation becomes more complicated when more bands are present. For this purpose we study the Uhlmann spectral flow of the Hofstadter model. Whereas in two band models, the amplitude spectrum can only be either gapped or gapless, in multi-band models, the amplitudes in general fall into groups where different groups are separated by a spectral gap, but levels within the same group evolve into each other (i.e. gapless). In such cases we take the sum of the phases of eigenvalues belonging to the same group as their collective Uhlmann phase (method 1 in the general discussion). Fig. 6.4 plots the Uhlmann spectral flow at four different temperatures with different gap structures in the amplitudes. As one can see, the gap structure depends sensitively on the temperature. A change in the gap structure indicates a topological transition. For $g$ amplitude gaps, there are $g + 1$ winding numbers but only $g$ of them are independent due to the restriction that their sum be zero. Much like the energy bands, there is a conservation of winding number when an amplitude gap emerges or collapses.

6.7 Discussion

While our primary interest is to formally extend the zero-temperature topological indices for pure states to mixed states at finite temperature, we note that it is experimentally relevant [20, 1, 81, 67]. In the context of band insulators, there already exists experimental technique [2] to measure Berry phase of pure electronic states in 1D. Such techniques can be used to measure the Uhlmann phase $\tilde{\gamma} = \text{Tr} M$ through the so-called purification procedure, where the amplitude matrix $W$ of a mixed state is mapped to a pure state in an enlarged system (the reduced density matrix of which is the mixed state). The Uhlmann phase $\tilde{\gamma}$ of the mixed state is identified with the Berry phase of the enlarged system, and is thus measurable. It is not a far stretch to envisage an adiabatic pumping of such 1D systems with a periodic external parameter, from which one can observe the winding of the
Figure 6.4: Uhlmann spectral flow of the Hofstadter model with \( p/q = 3/7 \). Top panels: amplitudes of eigenvalues. Central panels: collective phases (to be explained later). Bottom panels: schematics of the eigenvalue spectral flows; Empty boxes represent amplitudes of eigenvalues, and the color of their label corresponds to the color used in the top panels; Heavy horizontal lines indicate gaps in the amplitude spectrum, and boxes connected by thin lines indicate that the amplitudes they represent evolve into each other over \( k_x = 0 \rightarrow 2\pi \); Solid colored boxes on the right represent the corresponding group of connected levels, the winding number of the sum of the phases within each group is noted in the colored boxes, with the same color used in the central panels.
Uhlmann phase experimentally.

Chapter 7

Impurity scattering in Weyl Semimetals

A topological phase transition is usually associated with a quantum critical point, where the valence and conduction bands touch and exchange some topological number. In a 3D Weyl semimetal (WS), the critical point broadens into a phase throughout which one or more energy nodes persist and cannot be eliminated from bulk perturbation. The density of states (DOS) vanishes at these nodes. They appear as the ends of the so-called Fermi arc, i.e. the locus of gapless surface states interpolating between projections of Weyl nodes on the surface Brillouin zone.

The question we will address in this chapter is if the DOS suppression on the Weyl nodes can be lifted by resonance scattering of localized impurities. A localized impurity is defined as being restricted to a single unit cell but may affect electron’s internal degrees of freedom, for example a magnetic impurity scatters between different spin states. If the answer is affirmative, then a “zero mode” (a resonance at the nodal energy) can in principle be engineered which, lying in the energetic valley, will not be obscured by the bulk states, and is thus potentially useful in application. If the answer is negative, then there is a protection of the DOS suppression, the robustness of which might again serve some practical use. The question is also interesting from a technical point of view: edge/surface states of topological materials can be regarded as in-gap states induced by a particular type of defect — an edge or surface — in an otherwise translationally invariant
system. In WS, existence of Fermi arcs guarantees such in-gap states, thereby lifting the nodal DOS suppression on the surface. The question is then whether this warranty extends to localized impurities.

7.1 Introduction

The history of relativistic (Dirac) fermions in solid state band structures began with Wallace [69], who first considered a single layer of hexagonal graphite, i.e. graphene. It was however generally believed that such structures were intrinsically unstable and impractical to fabricate, but advances in materials preparation and experimental techniques have led to a surge of interest [14] and a new class of materials, known as Dirac materials [73, 26, 54, 72]. These materials have one or more symmetry-protected Dirac nodes where the density of state (DOS) vanishes and about which the energy dispersion is linear. The Dirac nodes are topological in nature, appearing as vortices or monopoles in the bulk Brillouin zone, and their presence typically requires some fine tuning such that the appearance of the Dirac structure marks a quantum phase transition between gapped phases. The properties of the gapped phases on either side of the transition differ in terms of their boundary behaviors, with the surface or edge spectra reflecting the topology of the bulk band structure. Familiar examples of systems with Dirac nodes include graphene and the class of materials known as topological insulators [14, 26, 54].

What distinguishes the Weyl semimetals within this framework is that the nodal structure exists not uniquely at a quantum critical point, but throughout an entire phase. In 3D, a Dirac point consists of two Weyl nodes of opposite chirality overlapping at the same point in $k$-space. While crystal symmetry may protect the two Weyl nodes from coupling [79, 71], thereby stabilizing the Dirac node, under a general perturbation they will be coupled and thereby open up a gap. In a system invariant under both time reversal ($T$) and inversion ($I$) symmetries, the nodal structure is at least four-fold degenerate (Kramers degeneracy), and the aforementioned separation requires the breaking of either $T$ or $I$ symmetry, or both. This can be achieved for example by introducing external field, magnetic
bulk impurities, electron interaction, etc. [10, 24, 23, 58]. If the perturbation which lifts the degeneracy is strong enough, a new gapless phase may result, which is the Weyl semimetal (WS) phase. The WS has been identified in several recent studies [43, 76, 77, 70, 28, 75, 36], and is a stable phase because further bulk perturbations can only shift the Weyl nodes without eliminating them. Recent work has also explored more exotic band structures involving Weyl fermions, for example their coexistence with quadratic (massive) fermions [48, 61], or the symmetry-enforced overlap of Weyl nodes of the same chirality [21]. It is known that band insulators may undergo a topological phase transition as the bulk gap collapses and re-opens again; in this sense, the WS – whose bulk gaps are closed – is an intermediation of two topologically distinct gapped phases [68], e.g. from a trivial insulator to a topological insulator [43], or a Chern insulator if $T$ is broken [10].

A minimal model of the WS was constructed by Burkov, Hook, and Balents (BHB) [11]. BHB considered a massless 3 + 1 dimensional four-component Dirac fermion model, initially with time-reversal and inversion symmetries. A $k \cdot p$ expansion about the $T$ and $I$-symmetric Brillouin zone center yields the four-component Hamiltonian $H^0(k) = \sum_{a=1}^{3} k_a \Gamma^a + m \Gamma^4$, written in terms of Dirac matrices. The sign of the mass $m$ tells if the bulk insulating state is normal or topological. BHB showed that adding a homogeneous $T$ and/or $I$ breaking term initially gaps out the Dirac spectrum, but that for sufficiently large symmetry breaking a WS phase appears, with Weyl nodes occurring at two distinct $k$ points. (In some cases, the two central bands touch along a circle in $k$-space.)

From the perspective of bulk-boundary correspondence, Weyl nodes may appear as the ends of the so-called Fermi arc [70], the locus of gapless surface states interpolating between the projections of Weyl nodes on the surface Brillouin zone. Such gapless modes participate in surface transport, with their multiplicity proportional to the arc length. This gives rise to an anomalous Hall effect (AHE) of the $T$ breaking WS, which recently has been shown to survive even when the Weyl nodes are subsequently gapped out by node-mixing scatterings, and is attributed to the persistence of chiral anomaly [82].

In this chapter, we investigate the effects of localized impurities on the bulk
electronic structure of the WS. In particular, we address the question of whether or not the DOS suppression at Weyl nodes in clean samples can be lifted via impurity scattering. Local impurities are modeled as $V = g\Lambda \delta(x)$ where $g \in \mathbb{R}$ is the coupling strength, $\delta(x)$ restricts the impurity to the site at $x = 0$, and $\Lambda$ is a matrix structure encoding its physical type, e.g., $\Lambda = I$ for scalar (chemical potential) impurity, and $\Lambda \propto \sigma_z$ for a magnetic impurity polarized along the $z$ direction. We will speak of the stability of an energy $\omega$ under the scattering of a $\Lambda$-type impurity in the following sense: if a resonance or bound state can be induced at $\omega$ by $\Lambda$ with some $g$, then $\omega$ is unstable with respect to $\Lambda$. Otherwise it is stable. Close to the nodal energy, the DOS vanishes as $\omega^2$. If the nodal energy is unstable, the resulting resonances will give rise to sharp peaks in the DOS which disrupt the pristine Dirac spectrum [9]. Bulk transport consequences for scalar impurities were considered in Refs. [11, 30, 24, 44, 7]. Ref. [44] studied the effect of rare regions in a dirty WS. Effects of scalar and magnetic impurities on the surface Dirac nodes of 3D topological insulators were studied in Ref. [8].

One might be tempted to draw intuition from the more familiar single-band problems and conclude that an impurity can induce resonance or bound state at arbitrary energy, given the freedom in choosing its strength $g$, making all energies unstable. We find that in the multiple-band case such as the WS, while this still holds for scalar impurities, it is not true in general. Instead, stability depends crucially on the type of impurity, which is mathematically classified by its commutation relation with the $\Gamma$ matrices in the local Green’s function. For some impurities, resonances and bound states are forbidden over a wide range of energies. Typically, an impurity is a foreign atom or local crystalline defect in an otherwise pristine material. Thus a realistic impurity potential should always involve a local scalar scattering component. If this scalar effect dominates the impurity, intragap resonances can be induced which will destabilize the Weyl node at a single particle level. If, on the other hand, the scattering is dominated by the resonance-forbidding components, then the Weyl node will remain.

This chapter is organized as follows: In Sec. 7.2, we present a general framework to address the existence of impurity resonances, and the dependence of their
energies on the impurity strength. In Sec. 7.3, we introduce a four-band tight binding lattice model of the WS in terms of the $\Gamma$ matrices, adapted from the continuum BHB model \cite{11}. Before analyzing the impurity effect in this lattice model, we first discuss in Sec. 7.4 the situation in the low energy theory, namely the original continuum BHB model, and show that a natural momentum cutoff around the Weyl nodes in such theories dismisses the important physics of a stabilized Weyl node. We then turn to a full lattice treatment: In Sec. 7.5, we apply the method developed in Sec. 7.2 to the WS model and classify impurities $\Lambda$ according to their effect on the electronic structure. We shall throughout the chapter restrict $\Lambda$ to being one of the sixteen $\Gamma$ matrices (including $\mathbb{I}$); linear combinations thereof can be analyzed in the same fashion but with more tedious algebra. We will first illustrate the method in Sec. 7.5.1 with the simpler case of a Dirac semimetal which is invariant under both $T$ and $I$. The (fine-tuned) Dirac node is shown to be unstable with $I$ even impurities, but stable with $I$ odd ones. We apply the same approach to WS with a symmetry breaking term $\eta \tilde{\Gamma}$ where $\tilde{\Gamma}$ is a matrix and $\eta$ is its strength. For a generic energy $\omega$, we find that its stability depends on both the impurity type $\Lambda$ and the symmetry breaking term $\eta \tilde{\Gamma}$. Sec. 7.5.2 discusses $I$ breaking WS (which may or may not break $T$) where $\tilde{\Gamma}$ anticommutes with $I$. In this case, the Weyl node energy is found to be stable for any $\Lambda$ that does not commute with the local Green’s function, but unstable if it commutes. Sec. 7.5.3 discusses the $I$-symmetric WS, in which $\tilde{\Gamma}$ necessarily breaks $T$. Again, impurities commuting with the local Green’s function will disrupt the Weyl node stability. Those that do not fully commute yield either a nodal energy stable over the full range of $\eta$, or a critical symmetry breaking amplitude $\eta_c$ – which is fully determined by parameters of the clean system – that splits the $\eta$ axis into two phases where the nodal energy is stable in one phase and unstable in the other. The critical amplitude $\eta_c$ is found to be related to a type of band inversion and indicates a phase transition in the gap of resonant impurity band structure, reminiscent of band inversions in topological/Chern insulators that are responsible for topological phase transitions. We conclude in Sec. 7.6.
7.2 Resonance criteria in a generic multi-band system

The effect of localized impurities can be studied in general using the standard $T$-matrix formalism [19, 3]. We briefly summarize the procedure below and establish notation. Given a Hamiltonian $H = H^0 + V$, its Green’s function is

$$G(z) \equiv (z - H)^{-1} = G^0 + G^0 T G^0,$$  \hspace{1cm} (7.1)

where $z \in \mathbb{C}$ is the complex frequency, $G^0(z) \equiv (z - H^0)^{-1}$ is the Green’s function of $H^0$, and $T = V(\mathbb{I} - G^0 V)^{-1}$ is the $T$-matrix. Assume the impurity potential $V$ is localized at a spatial point $r = 0$: $V_{rr'} = g \Lambda \delta_{r,0} \delta_{r',0}$, where $g$ is the potential strength and $\Lambda$ is a dimensionless matrix whose rank is equal to the number of bands. The Green’s function connecting $r$ and $r'$ is $G_{rr'} = G^0_{rr'} + G^0_{r0} T_{00} G^0_{0r'}$, and $T_{00}(z) = [g^{-1} \Lambda^{-1} - G^0_{00}(z)]^{-1}$ is the only nontrivial block of the $T$-matrix. For translationally invariant systems, the local Green’s function is $G^0_{rr} = G^0_{00}(z) = \frac{1}{N} \sum_k (z - H^0_k)^{-1}$ where $H^0_k$ is the Fourier transform of $H^0$ and $N$ is the number of $k$ points, i.e. the number of unit cells of the crystal. The corresponding local density of states (LDOS) at site $r$ with energy $\omega$ is $\rho_r(\omega) = -\text{Im} \text{Tr} G_{rr}(\omega + i0^+)$. Bound states and resonances are consequences of the energy spectrum reconstruction induced by impurities. Before taking the thermodynamic limit, eigenvalues of $H$ are poles of $T(\omega)$ on the real $\omega$ axis, i.e. the zeros of $T^{-1}(\omega) = g^{-1} \Lambda^{-1} - G^0_{00}(\omega)$. Upon tuning of the impurity strength $g^{-1}$, each pole will trace out a curve in the $(\omega, g^{-1})$ plane. Note that $g^{-1} = \pm \infty$ are identified as $g = 0$, and $g = \pm \infty$ as $g^{-1} = 0$. Thus one may consider an adiabatic cycle in which $g^{-1} : -\infty \to 0 \to +\infty$. After a full cycle, the poles must collectively recover their initial positions, which are the set of clean states at $g^{-1} = \pm \infty$. Individual poles may either stick close to one clean state, or migrate between different ones. An example is shown in Figs. 7.1a and 7.1b for the spectral evolution of a graphene sheet of $6 \times 6$ and $10 \times 10$ unit cells, respectively, in the presence of a scalar impurity.
Figure 7.1 (next page): Various ways to identify bound states and resonances for a scalar impurity $V = g \mathbb{I}$ in graphene. Clean graphene is modeled as $H^0(k) = \begin{pmatrix} 0 & \gamma_k \\ \gamma_k^* & 0 \end{pmatrix}$ where $\gamma_k = 1 + e^{-ik_1} + e^{-ik_2}$ and $k_i = k \cdot a_i$, with $a_i$ ($i = 1, 2$) being the two primitive direct lattice vectors. Panels (a) and (b) show the spectral evolution of $H = H^0 + g \mathbb{I}$ (solid and dotted gray curves) with impurity strength $g$, for different lattice sizes. These are solved as zeros of the eigenvalues of the $2 \times 2$ matrix $T^{-1}(\omega, g) = g^{-1} \mathbb{I} - G_{00}^0(\omega)$, with the eigenvalue index encoded by different line color/types. Migration of such levels between different clean states ($g^{-1} = \pm \infty$ limit) constitutes a resonance (inside bulk bands) or a bound state (outside bulk bands), and can be extracted as zeros of $\mathcal{T}^{-1}(\omega, g)$, the Hermitian part of $T^{-1}$, shown as dashed green curves, with solid circular blue points overlaying on the branch corresponding to the solid spectral flow and red empty square on the branch corresponding to the dotted spectral flow (see text). The discontinuity in the green curves at $\omega = \pm 1$ is concomitant with the Van Hove singularity in the DOS (not plotted). (c) shows the phase shift $\arg \det T(\omega + i0^+, g)$ in the thermodynamic limit, where $\pm \frac{\pi}{2}$ (heaviest red/blue) could be interpreted as resonance or bound state, see text. (d) plots the norm of retarded $T$ matrix, which can be used to distinguish between resonance and anti-resonance that is hard to tell from (c), the former corresponding to the dark feature and the latter suppressed in such a plot. The dashed green curves in (c) and (d) are the same as those in (a) and (b).
A point \((\omega, g^{-1})\) on the spectral evolution curves outside the bands of the clean system represents a bound state of energy \(\omega\) induced by an impurity of strength \(g\). For those inside the clean bands, one has to distinguish between poles very close to – hence mere perturbations of – a clean state, and those in the middle of a migration. The latter, like bound states, are manifestations of the impurity effect. They differ in that an increase in system size \(N\) has little influence on the bound states, but will split spectral lines inside the clean bands to accommodate
newly created clean states; see Fig. 7.1a and 7.1b for this lattice size effect. What remains unchanged when increasing $N$ is the trend of rapid pole migration.

An effective way to extract the locus of such rapid migrations, which we identify as resonances, is to find the zeros of $\mathcal{T}^{-1}(\omega, g^{-1}) \equiv g^{-1} \Lambda^{-1} - \mathcal{G}_{00}^{0}(\omega)$, where $\mathcal{G}_{00}^{0}(\omega) \equiv \frac{1}{2}(\mathcal{G}_{00}^{0}(\omega + i\epsilon) + [\mathcal{G}_{00}^{0}(\omega + i\epsilon)]^\dagger)$ is the Hermitian part of the retarded local Green’s function, where the imaginary part $\epsilon$ is taken to be greater than the energy spacing between consecutive bulk levels. In so doing, divergences in $T$ originating from poles of $\mathcal{G}_{00}^{0}(\omega)$ itself are eliminated from the zeros of $\mathcal{T}^{-1}(\omega)$, leaving only those caused by the aforementioned spectral migrations. The zeros of $\mathcal{T}^{-1}(\omega)$ are shown for the case of graphene as green curves in Figs. 7.1a and 7.1b. In these figures, they sit close to the inflection points of the spectral curves, where one might say the pole migration is most rapid.

In the thermodynamic limit $N \to \infty$, the Hermitian matrix $\mathcal{T}^{-1}$ remains well defined. Since $\epsilon \to 0^+$, $\mathcal{T}^{-1}$ and $T^{-1}$ are identical outside bulk bands, the zeros of $\mathcal{T}^{-1}$ can be used to identify both bound states and resonances. Callaway showed [12] that the phase shift at $\omega$, viz. $\delta(\omega) \equiv \text{arg det } T(\omega + i0^+)$, equals $\pi \times \Delta N(\omega)$ where $\Delta N(\omega)$ is the difference in the total number of states of $H$ and of $H^0$ below $\omega$. One definition of resonance in this context is for the phase shift to be $\pm \pi/2$, viz. $\text{Re det } T = 0$, the reason being that the number of extra states is a half-odd-integer, which represents the “center” of the process in which one extra state is gained or lost. The former is called a resonance and the latter an anti-resonance. Over the full range of $\omega$ they must balance each other if introduction of an impurity does not change the total number of states, a form of the Friedel sum rule. This is related to our resonance criteria in that the zero of $\mathcal{T}^{-1}$ is the center of a spectral migration – a spectral line migrating past $\omega$, by definition, contributes $|\Delta N(\omega)| = 1$ to the total number of states below $\omega$. The difference between our criteria and the phase shift picture is that there are in general $N_B$ branches of spectral evolutions at any $\omega$ where $N_B$ is the number of bands, e.g. $N_B = 2$ in the graphene example shown in Fig. 7.1. Our criteria is essentially to track the increment/decrement contributed by any single branch, whereas $\text{Re det } T = 0$ takes into account all branches. In any case, the difference is consequential only in identifying the location of the anti-
resonances. For comparison, we plot the phase shift (color map) together with the zeros of $T^{-1}$ (dotted green line) for a graphene impurity in Fig. 7.1c.

The distinction between resonance and anti-resonance is not immediately apparent from the phase shift plot. It relies on the sign of $s = \partial \delta / \partial \omega$: $s > 0$ is a resonance and $s < 0$ an anti-resonance. Furthermore, in cases where two bound states/resonances are close together – or even degenerate as can happen in Dirac semimetals to be discussed in Sec. 7.5.1 – the phase shift will experience a $2\pi$ change over a small $\omega$ window, which is equivalent to zero numerically and hence hard to resolve. A more transparent way is to plot the matrix norm $||T_{00}(\omega + i0^+, g^{-1})|| = \sqrt{\sum_a |\lambda_a(\omega + i0^+, g^{-1})|^2}$ over the $(\omega, g^{-1})$ parameter space, where $\lambda_a(z, g^{-1})$ are (complex) eigenvalues of $T_{00}(z, g^{-1})$. If $||T_{00}(\omega + i0^+, g^{-1})||$ is large, then the DOS is in general enhanced, and one obtains a resonance. This is shown for the graphene example in Fig. 7.1d.

To facilitate analytical treatment, we will henceforth use the $T^{-1}$ approach and make no distinctions among resonance, anti-resonance, and bound state. All of them will simply be referred to as “resonance”. As needed we will also exhibit $||T(\omega + i0^+)||$ plots, where anti-resonances are suppressed, as a check.

Consider next the existence of a resonance at an arbitrary point in $(\omega, g^{-1})$ space, i.e. the condition for at least one eigenvalue of $T^{-1}(\omega, g^{-1})$ to be zero. If $g$ is allowed to be complex, then there are as many solutions of $g$ at a given $\omega$ as the number of bands: $g^{-1} = u_a(\omega)$, where $\{u_a(\omega)\}$ are the eigenvalues of $G_{00}^0(\omega)\Lambda$. However, only real $g$ is physical. Thus existence of resonance demands at least one eigenvalue of $G_{00}^0\Lambda$ to be real. The required coupling strength is $g = 1/u_a(\omega)$. An immediate corollary is that impurities with $[\Lambda, G_{00}^0] = 0$ can induce a resonance at arbitrary energy for some $g$, because product of commuting Hermitian matrices has real eigenvalues. Single-band problems fall in this category ($\Lambda = I$).

### 7.3 Weyl semimetal models

We now apply the method described above to lattice systems adapted from the continuum models of BHB [11]. The $k$-space Hamiltonian in the $\Gamma$-matrix
Table 7.1: Symmetry of $\Gamma$ matrices. $T$ and $l$ stand for time-reversal and inversion, respectively. A plus sign indicates that the associated $\Gamma$ matrices commute with the corresponding symmetry operation, and a minus sign indicates anticommutation.

<table>
<thead>
<tr>
<th>matrices</th>
<th>$T$</th>
<th>$l$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbb{I}, \Gamma^4$</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>$\Gamma^1, \Gamma^2, \Gamma^3, \Gamma^5$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$\Gamma^{12}, \Gamma^{13}, \Gamma^{23}, \Gamma^{15}, \Gamma^{25}, \Gamma^{35}$</td>
<td>$-$</td>
<td>+</td>
</tr>
<tr>
<td>$\Gamma^{14}, \Gamma^{24}, \Gamma^{34}, \Gamma^{45}$</td>
<td>+</td>
<td>$-$</td>
</tr>
</tbody>
</table>

The basis is

$$H^0(k) = \xi(k)\mathbb{I} + \sum_{i=1}^{3} d_i(k)\Gamma^i + m(k)\Gamma^4 + \eta\tilde{\Gamma}$$  \hspace{1cm} (7.2)

where the $k$-dependent coefficients are taken to be

$$\xi(k) = -2t \sum_{i=1}^{3} \cos k_i - \varepsilon_0, \quad d_i(k) = -2t_1 \sin k_i, \quad m(k) = -4t' \sum_{i=1}^{3} (1 - \cos k_i) - \lambda,$$

and $\eta$ is $k$-independent for simplicity. $\tilde{\Gamma}$ is a $\Gamma$ matrix which breaks time-reversal ($T$) and/or inversion ($l$) symmetry to be defined later. The following $\Gamma$ matrix convention is used:

$$\Gamma^i = \tau^x \otimes \sigma^i \quad (i = 1, 2, 3), \quad \Gamma^4 = \tau^y \otimes \mathbb{I}, \quad \Gamma^5 = -\tau^y \otimes \mathbb{I}, \quad \Gamma^{\mu\nu} = i[\Gamma^\mu, \Gamma^\nu]/2$$  \hspace{1cm} (7.3)

where $\tau^i$ and $\sigma^i$ are two sets of Pauli matrices acting on the orbital and spin degrees of freedom, respectively. In this model, $t$ and $t'$ are intra-orbital hoppings, while $t_1$ is the (spin-mixing) hopping between different orbitals. Different conventions of $\Gamma$ matrices may have different physical interpretations: for example, in the convention above, $\Gamma^{12} = -\mathbb{I} \otimes \sigma^z$ represent a magnetic field in the $z$ direction, but in other conventions it may also have orbital effects. Results obtained below are independent of the convention used.

BHB showed how the emergence of stable point or line nodes in the spectrum of $H^0$ beyond a critical perturbation strength depends on the symmetry of the perturbation [11]. Define time reversal as $T = KR$ where $K$ is complex conjugation and $R = \mathbb{I} \otimes i\sigma^y$, and inversion as $l = \Gamma^4$. Symmetry properties of all $\Gamma$
matrices can be found in Table 7.1. If \( \eta = 0 \), then the model is both \( T \) and \( I \)-symmetric. In this case, fine-tuning \( \lambda = 0 \) creates a Dirac node at \( k = 0 \) where the four bands converge to the energy \( E = -6t - \varepsilon_0 \). While the Dirac point is gapped out by nonzero \( \lambda \), and is therefore unstable in the \( T \) and \( I \)-symmetric system, it splits into two Weyl nodes if either \( T \) or \( I \) is broken by \( \tilde{\Gamma} \) [11, 43]. In this case the nodal structure survives in a range of parameters and constitutes a stable nodal phase, \( \text{i.e.} \) the Weyl semimetal (WS) phase.

In BHB’s language, \( H^0 \) with \( \eta = 0 \) is the unperturbed Hamiltonian, and the \( \eta \tilde{\Gamma} \) term is the symmetry-breaking perturbation. Although the Weyl nodes and hence the semimetal phase are stable under such \textit{homogeneous} bulk perturbations, the characteristic suppression of DOS at the nodal energy in the WS phase, may be destroyed by another type of perturbation – namely localized impurity potentials – if resonances can be induced at the nodal energy. Hereafter, we shall refer to the full \( H^0 \) of Eq. 7.2, including the homogeneous term \( \eta \tilde{\Gamma} \), as the unperturbed Hamiltonian, and regard the local impurity potential as the perturbation.

The model of Eq. 7.2 is lattice-based, which entails a specific cutoff structure, and hence not generic like the BHB Hamiltonian \( H_{\text{BHB}} = \sum_{i=1}^{3} k_i \Gamma_i + m \Gamma_4 + \eta \tilde{\Gamma} \). It will be instructive to first look at the effect of scattering from the more universal low energy states living in the vicinity of Weyl nodes, which we shall analyze in Sec. 7.4 for the BHB model. However, as we shall see there, the very act of taking a momentum cutoff will leave out the possibility of a stable Weyl node. This is not surprising because spatially localized impurities are homogeneous in the momentum space and inevitably scatter high momentum states. Thus a lattice treatment is necessary. For the lattice theory, we will first discuss the impurity effect in the \( T \) and \( I \)-symmetric system (\( \eta = 0 \)). While it does not yield a WS phase, it is simple enough to be used as a demonstration of the general framework outlined in Sec. 7.2. Then we will move on to unperturbed systems with \( T \) and/or \( I \) broken (\( \eta \neq 0 \)) where a WS phase does exist. Since the Hermitian part of the local unperturbed Green’s function, \( G_{00}^0 \), is of central importance to the impurity classification, we will first classify the \textit{unperturbed} system according to the type of \( \Gamma \) matrices appearing in \( G_{00}^0 \), and then for each of them, classify the impurities
7.4 Impurity scattering in the low energy theory

In this section we focus on the low energy theory described by the following BHB Hamiltonian,

$$H_{\text{BHB}} = \sum_{i=1}^{3} k_i \Gamma_i + m \Gamma^4 + \eta \tilde{\Gamma}.$$  \hspace{1cm} (7.4)

Consider for example $\tilde{\Gamma} = \Gamma_{21}$ and $\Gamma_{35}$. The eigenvalues of $H_{\text{BHB}}$ are $\pm E_s$ where

$$E_s = \begin{cases} \sqrt{k_x^2 + k_y^2 + (\eta + s \sqrt{k_x^2 + m^2})^2}, & \tilde{\Gamma} = \Gamma_{21} \\ \sqrt{k_z^2 + (\eta + s \sqrt{k_x^2 + k_y^2 + m^2})^2}, & \tilde{\Gamma} = \Gamma_{35} \end{cases}, \quad s = \pm 1.$$  \hspace{1cm} (7.5)

In both cases the scale of nodal momentum is $\Delta = \sqrt{\eta^2 - m^2}$. For $\eta > m > 0$, $\tilde{\Gamma} = \Gamma_{21}$ generates two point nodes at $k = (0, 0, \pm \Delta)$, whereas $\tilde{\Gamma} = \Gamma_{35}$ generates a nodal line at $k = (\Delta \cos \phi, \Delta \sin \phi, 0)$. In the vicinity of the Weyl nodes, the $s = -1$ bands have linear dispersion, whereas the $s = +1$ pair is gapped, $E_+ = 2\eta + O(q^2/\eta^2)$.

If we are looking for low energy resonances, $\omega \sim 0$, then since the Green’s function is weighted by $1/(\omega - E)$, it is reasonable to focus on the scattering of low energy states by (i) projecting onto the $s = -1$ bands, and (ii) adopting a momentum cutoff such that only momenta within a distance $Q$ from the nodes are considered – a sphere around point nodes and a tube around nodal line – with $\omega \ll Q \ll \eta$. In this approximation the local Green’s function for both $\tilde{\Gamma}$s have the following form,

$$G_{00}^0(\omega) = a(\omega) \left( I - \frac{m}{\eta} \Gamma^4 \tilde{\Gamma} \right),$$  \hspace{1cm} (7.6)

with

$$a(\omega) = \begin{cases} \frac{Q \omega \eta}{4 \pi^2 m} \log \frac{\eta - m}{\eta + m} - \frac{i \eta}{4 \pi \Delta} \omega^2, & (\tilde{\Gamma} = \Gamma_{21}) \\ \frac{\Delta}{2} R(m/\eta) \omega - \frac{i \eta}{8} |\omega|, & (\tilde{\Gamma} = \Gamma_{35}) \end{cases}.$$  \hspace{1cm} (7.7)
see D.2 for derivation and Eq. D.78 for the expression of $R(m/\eta)$. The hermitian $G^0_{00}(\omega)$ is obtained by taking the real part of $a(\omega)$.

Let us now analyze the resonance condition. The $T^{-1}$ matrix is

$$T_{00}^{-1}(\omega) = g^{-1} \Lambda + a_r(\omega) \frac{m}{\eta} \Gamma^4 \tilde{\Gamma} - a_r(\omega) \mathbb{I}$$

(7.8)

where $a_r$ is the real part of $a$. The impurity $\Lambda$ either commutes or anticommutes with $\Gamma^4 \tilde{\Gamma}$ since the latter is itself one of the sixteen $\Gamma$ matrices. If they commute, then $\det T_{00}^{-1} = 0$ yields real solutions for $g^{-1}$, i.e. resonance could be induced and $\omega \sim 0$ is unstable. If, on the other hand, $\Lambda$ and $\Gamma^4 \tilde{\Gamma}$ anticommute, then $\det T_{00}^{-1} = 0$ gives

$$g^{-1} = \pm |a_r(\omega)| \sqrt{1 - \frac{m^2}{\eta^2}}.$$  

(7.9)

Now, by $m^2 + \Delta^2 = \eta^2$, one obtains $g^{-1} = \pm |a_r(\omega)| \Delta/\eta$, which is still real, i.e., $\omega \sim 0$ is unstable for anticommuting $\Lambda$. Thus in the cutoff scheme adopted here, $\omega \sim 0$ is unstable regardless of the type of impurity. Note however that in the case of anticommuting ones, the impurity strength $g$ always comes in $\pm$ pairs, whereas in the commuting case it might not.

An essential difference between the commuting and anticommuting impurities is that in the latter case, the solution for $g^{-1}$ contains a square root. When higher momentum states are considered, the argument of the square root may become negative, and stabilize the nodal energy.

To see this, note that without the low energy restriction, the local Green’s function of the BHB Hamiltonian with $\tilde{\Gamma} = \Gamma^{21}$ and $\Gamma^{35}$ has the form (see Eq. D.20 in section D.1)

$$G^0_{00}(\omega) = a(\omega) \mathbb{I} + b_1(\omega) \Gamma^4 + b_2(\omega) \tilde{\Gamma} + b_3(\omega) \Gamma^4 \tilde{\Gamma}$$

(7.10)
where the coefficients are

\[ a(\omega) = \frac{\omega}{2} \left( \frac{1}{\omega^2 - E_+^2} + \frac{1}{\omega^2 - E_-^2} \right), \]  

(7.11)

\[ b_1(\omega) = \frac{m}{2} \left( \frac{1}{\omega^2 - E_+^2} + \frac{1}{\omega^2 - E_-^2} + \frac{4\eta^2}{(\omega^2 - E_+^2)(\omega^2 - E_-^2)} \right), \]  

(7.12)

\[ b_2(\omega) = \frac{\eta}{2} \left( \frac{1}{\omega^2 - E_+^2} + \frac{1}{\omega^2 - E_-^2} + \frac{4d_\perp^2}{(\omega^2 - E_+^2)(\omega^2 - E_-^2)} \right), \]  

(7.13)

\[ b_3(\omega) = 2\eta\omega m \left( \frac{1}{(\omega^2 - E_+^2)(\omega^2 - E_-^2)} \right), \]  

(7.14)

in which \( \langle \cdots \rangle \) denotes the \( \mathbf{k} \)-space integral (for continuum) or sum (for lattice), and

\[ d_\perp^2 = \begin{cases} 
  k_z^2 + m^2 & (\tilde{\Gamma} = \Gamma^{21}) \\
  k_x^2 + k_y^2 + m^2 & (\tilde{\Gamma} = \Gamma^{35}) 
\end{cases} \]  

(7.15)

In the low energy approximation, \( b_1 \) and \( b_2 \) vanish due to the momentum cutoff. This is because at the Weyl nodes, \( E_+^2 = d_\perp^2 = 4\eta^2 \), thus in the low energy approximation, the second and third terms inside \( \langle \cdots \rangle \) of both \( b_1 \) and \( b_2 \) cancel, yielding

\[ b_1(\omega) \approx \frac{m}{2(\omega^2 - E_+^2)} \frac{V_{\text{cutoff}}}{V_{\text{BZ}}}, \quad b_2(\omega) \approx \frac{\eta}{2(\omega^2 - E_+^2)} \frac{V_{\text{cutoff}}}{V_{\text{BZ}}}, \]  

(7.16)

where \( V_{\text{cutoff}}/V_{\text{BZ}} \) is the ratio between the volume within the momentum cutoff and that of the first Brillouin zone. This volume ratio comes from the evaluation of \( \langle 1 \rangle \), and is of the order \((Q/2\pi)^{\text{codim}} \sim (Q/\eta)^{\text{codim}} \to 0\) in the low energy approximation, with \( \text{codim} \) being the codimension of the Weyl node, which is 3 for a point node and 2 for a nodal line. When higher momentum states are included, \( b_1 \) and \( b_2 \) will no longer be suppressed, and will change the argument under the square root, possibly making it negative and stabilizing the nodal energy. A more careful analysis necessitates a lattice treatment, which is what we shall do in the rest of the chapter.
7.5 Classification of impurity potentials

7.5.1 \textbf{T} and \textbf{I}-symmetric $H^0(k)$

We now return to the lattice model of Sec. 7.3. To illustrate the resonance criteria of Sec. 7.2, we first consider the case with $\eta = 0$. Inverting Eq. 7.2 yields the unperturbed $k$-space Green’s function,

$$G^0(\omega, k) = \frac{[\omega - \xi(k)] I + \sum_{i=1}^{3} d_i(k) \Gamma_i + m(k) \Gamma^4}{[\omega - \xi(k)]^2 - \sum_{i=1}^{3} |d_i(k)|^2 - m^2(k)}.$$  \hfill (7.17)

The $\Gamma_i$ ($i = 1, 2, 3$) terms will vanish after summation over $k$ due to the oddness of $d_i(k)$, as required by inversion symmetry. The local Green’s function is thus

$$G^0_{00}(\omega) = \frac{1}{N} \sum_k G^0(\omega, k) = a(\omega) I + b(\omega) \Gamma^4,$$  \hfill (7.18)

where $N$ is the number of $k$ points,

$$a(\omega) = \frac{1}{N} \sum_k \frac{\omega - \xi(k)}{D(\omega, k)} , \quad b(\omega) = \frac{1}{N} \sum_k \frac{m(k)}{D(\omega, k)} ,$$  \hfill (7.19)

and

$$D(\omega, k) = [\omega - \xi(k)]^2 - \sum_{i=1}^{3} d_i^2(k) - m^2(k).$$  \hfill (7.20)

As discussed in Sec. 7.2, the existence of a resonance depends on whether or not the eigenvalues of $T^{-1}_{00}$, \textit{i.e.} the Hermitian part of the inverse local $T$-matrix, can be zero. Since the only $\Gamma$ matrix in the $G^0_{00}$ decomposition is $\Gamma^4 = I$, there are only two classes of impurities according to their inversion property:

\textbf{Inversion-even impurity}

In this class we have $\Lambda = I, \Gamma^4, \text{or } \Gamma^{\mu\nu}$ ($\mu, \nu = 1, 2, 3, 5$). Since they all commute with $G^0_{00}$, a resonance can be induced at arbitrary energy, \textit{i.e.} a solution of $\det T^{-1}_{00} = 0$ exists for real $g$. To illustrate, we solve for $g(\omega)$, the value of $g$ which produces a resonance at energy $\omega$, for all three cases:

1. $\Lambda = I$: This is a scalar impurity, and

$$T_{00}^{-1}(\omega) = g^{-1} - G^0_{00}(\omega) = [g^{-1} - a(\omega)] I - b(\omega) \Gamma^4.$$  \hfill (7.21)
The principal values of $a$ and $b$ are implicitly taken. Setting the LHS to zero yields

$$g^{-1}(\omega) = a(\omega) \pm b(\omega) .$$

(7.22)

These are shown as the light blue dashed lines in Fig. 7.2(b).

(2) $\Lambda = \Gamma^4$: This impurity flips the sign of the inversion-odd component, yielding

$$T_{00}^{-1}(\omega) = -a(\omega) \mathbb{I} + [g^{-1} - b(\omega)] \Gamma^4 .$$

(7.23)

The resonance condition is thus

$$g^{-1}(\omega) = b(\omega) \pm a(\omega) .$$

(7.24)

(3) $\Lambda = \Gamma^{\mu\nu}$ ($\mu, \nu = 1, 2, 3, 5$): This includes for example the magnetic impurities, $\Gamma^{12} = -\mathbb{I} \otimes \sigma^z$, etc, and

$$T_{00}^{-1}(\omega) = g^{-1} \Lambda - a(\omega) \mathbb{I} - b(\omega) \Gamma^4 .$$

(7.25)

The eigenvalues of $T_{00}^{-1}$ are obtained by replacing $\Lambda$ and $\Gamma^4$ on the RHS each with uncorrelated $\pm 1$ (since they can be simultaneously diagonalized). Setting these eigenvalues to zero yields

$$g^{-1}(\omega) = \pm a(\omega) \pm b(\omega) .$$

(7.26)

These are shown as dash and dash-dot lines in Fig. 7.2(b).

**Inversion-odd impurity**

In this class we have $\Lambda = \Gamma^\mu$ or $\Gamma^{4\mu}$ with $\mu = 1, 2, 3, 5$. The inverse $T$ matrix is

$$T_{00}^{-1}(\omega) = g^{-1} \Lambda - a(\omega) \mathbb{I} - b(\omega) \Gamma^4 .$$

(7.27)

Rearranging and squaring, one gets $(T_{00}^{-1} + a\mathbb{I})^2 = g^{-2} + b^2$. The cross term of $g^{-1}$ and $b$ vanishes because $\{\Lambda, \Gamma^4\} = 0$. Setting $T_{00}^{-1} = 0$ then yields

$$g^{-1}(\omega) = \pm \sqrt{a^2(\omega) - b^2(\omega)} .$$

(7.28)

The results are plotted as dashed lines in Fig. 7.2 (a). Thus inversion-odd impurities cannot induce any resonance in the range of energy $\omega$ for which

$$|a(\omega)| < |b(\omega)| .$$

(7.29)
Figure 7.2: Resonance condition for $T$ and $I$-symmetric Dirac Semimetals. Color map: $\log ||T_{00}(\omega + i\epsilon)||$ for $T$ and $I$-symmetric case ($\eta = 0$). Darker color corresponds to stronger impurity effect. Unperturbed DOS is shown at the bottom. Colored lines interpolating the dark curves are obtained by replacing $G^0_{00}$ with $G^0_{00}$ and computing the zeros of $g^{-1}\Lambda - G^0_{00}(\omega)$ in the $g - \omega$ plane (see text). For $\Lambda = \Gamma^1$ (a), the Dirac node is stable. Panel (b) shows results for $\Lambda = \Gamma^{12}$ (all curves) and for $\Lambda = \mathbb{I}$ (blue dashed curves only); the Dirac node is unstable. Parameters used are $t = 0.05, t_1 = -0.5, t' = -0.25, \lambda = 0, \varepsilon_0 = -0.3$, and lattice size $N_x = N_y = N_z = 50$. Spectral broadening $\epsilon$ is set to 0.05.
Band center approximation (BCA)

To understand the general trend of the resonance solutions \( g(\omega) \) and get a sense of the stability region, it is useful to obtain an approximation for the expansion coefficients \( a(\omega) \) and \( b(\omega) \). To this end we introduce the band center approximation (BCA): A generic \( k \)-space Hamiltonian \( H(k) \) can be written as

\[
H(k) = H_{00} + \delta H_k
\]

where \( H_{00} = \langle H(k) \rangle \) is the local Hamiltonian and \( \langle \cdots \rangle \) denotes \( k \)-space averaging. The eigenvalues of \( H_{00} \) can be thought of as some sort of average energy of the bands of \( H(k) \) (band centers). Let

\[
\bar{G}(\omega) = (\omega - H_{00})^{-1}
\]

then the local Green’s function is

\[
G_{00} = \langle (\omega - H_{00} - \delta H_k)^{-1} \rangle
= \bar{G} + \bar{G} \langle \delta H_k \rangle \bar{G} + \bar{G} \langle \delta H_k \bar{G} \delta H_k \rangle \bar{G} + \cdots
= \bar{G} \left\{ 1 + \mathcal{O}((\delta H_k \bar{G})^2) \right\}
\]

where we have used \( \langle \delta H_k \rangle = 0 \). The BCA amounts to replacing the local Green’s function, \( G_{00} \), with the Green’s function of the local Hamiltonian, \( \bar{G} \).

Eq. 7.31 is an expansion in powers of \( \delta H_k / (\omega - H_{00}) \), where the numerator is roughly the band width, and the denominator is the distance from \( \omega \) to the band centers. The BCA works well if the distance of \( \omega \) from some band center, say that of band \( A \), is greater than \( A \)’s bandwidth. Note that such an \( \omega \), although outside band \( A \), may well be inside another band, say band \( B \). From the BCA point of view, a resonance at some \( \omega \) inside band \( B \) is actually a consequence of the coherent superposition of states mainly in some other band (\( A \)). The multiple-band scenario is to the benefit of the BCA.

Applying BCA to the \( \eta = 0 \) model here, one finds from Eq. 7.2 that \( H_{00} = \alpha I + \beta \Gamma^4 \) where \( \alpha = -\varepsilon_0 \) and \( \beta = -12t' - \lambda \), hence

\[
\bar{G} = \bar{a} I + \bar{b} \Gamma^4 \quad , \quad \bar{a} = \frac{\omega - \alpha}{(\omega - \alpha)^2 - \beta^2} \quad , \quad \bar{b} = \frac{\beta}{(\omega - \alpha)^2 - \beta^2} .
\]

\( \bar{G} \) and \( G_{00} \) have the same form of decomposition. This will prove useful in the more complicated situations where \( \bar{\Gamma} \) is present.
**Table 7.2**: Impurity classification for $I$ breaking Weyl material. The symmetry breaking term in $H^0(k)$ is $\Gamma = \Gamma^{\mu 4}$ ($T$ even) or $\Gamma^{\mu}$ ($T$ odd) where $\mu \in \{1, 2, 3, 5\}$. The second column indicates commutation (+) or anticommutation (−) of the impurity matrix $\Lambda$ with $\Gamma^4$ and $\tilde{\Gamma}$, respectively. Elements in each class are enumerated in the first column: if the cell has two sub-cells, the left one corresponds to $\tilde{\Gamma} = \Gamma^{\mu 4}$ and right one $\tilde{\Gamma} = \Gamma^{\mu}$; otherwise the enumeration is identical for both $\tilde{\Gamma}$. The value of $g$ at which $T^0_{\mu \nu}$ has a zero eigenvalue is listed in the third column, and the condition for it to be real (the resonance condition) is shown in the fourth column. The fifth column shows the resonance condition as given by the band center approximation, which are simple expressions in terms of the Hamiltonian parameters. Note that the values of the Green’s function coefficients $a(\omega)$ and $b_i(\omega)$ depend on the choice of $\tilde{\Gamma}$ that breaks $I$, but the BCA conditions are independent of $\tilde{\Gamma}$. The stability of Weyl nodes, if they exist, is listed in the last column (S for stable, U for unstable).

| $\Lambda$ ($\Gamma^{\mu 4} | \Gamma^{\mu}$) class | $g^{-1}$ | resonance | resonance (BCA) | node stability |
|----------------|----------|------------|-----------------|----------------|
| $\Gamma^{4}$ $\Gamma^{\mu \bar{\nu}}$ (+, −) | $b_1 \pm \sqrt{a_1^2 - b_2^2}$ | $|a| > |b_2|$ | $|\omega - \alpha| > |\eta|$ | S |
| $\Gamma^{\mu 4}$ $\Gamma^{\mu}$ (−, +) | $b_2 \pm \sqrt{a_1^2 - b_1^2}$ | $|a| > |b_1|$ | $|\omega - \alpha| > |\beta|$ | |
| $\Gamma^{\mu}$ $\Gamma^{\bar{\mu} 4}$ (−, −) | $\pm \sqrt{a_1^2 - b_1^2}$ $a^2 > b_1^2 + b_2^2$ $(\omega - \alpha)^2 > \beta^2 + \eta^2$ |
| $\Gamma^{\bar{\mu} 4}$ $\Gamma^{\mu 4}$ (−, +) | $a \pm \sqrt{b_1^2 + b_2^2}$ | any $\omega$ | any $\omega$ | U |

$\mu \in \{1, 2, 3, 5\}$ (fixed in $H^0$), $\bar{\mu}, \bar{\nu} \in \{1, 2, 3, 5\} \setminus \{\mu\}$. $\alpha = -\varepsilon_0$, $\beta = -12t' - \lambda$. 

\[a^2 > b_1^2 + b_2^2\]
For the scalar potential $\Lambda = \mathbb{I}$, the BCA resonance solution $g^{-1} \simeq \bar{a}(\omega) \pm \bar{b}(\omega) = [\omega - (\alpha \pm \beta)]^{-1}$ resembles two hyperbolae centered around the band centers $\alpha \pm \beta$. They can be identified qualitatively from the light blue dashed curves in Fig. 7.2(b), although numerically the two branches of each hyperbola, instead of being divergent, are connected around their respective band centers due to higher order effects in Eq. 7.31. For $l$ odd impurities, such as the magnetic impurity $\Lambda = \Gamma_{12}$, the stability condition Eq. 7.29 implies $\alpha - |\beta| < \omega < \alpha + |\beta|$, i.e., stable energy $\omega$ is bounded by the two band centers, as can be seen from Fig. 7.2 (a). This region in particular includes the (fine-tuned) Dirac point or the central gap. We thus conclude that the Dirac node is generically stable for $l$ odd impurities and unstable for $l$ even impurities.

### 7.5.2 $H_0^0(k)$ with $l$ breaking $\tilde{\Gamma}$

We now consider the case where $\tilde{\Gamma}$ breaks inversion. In the BHB scheme, this can be realized for example by applying a voltage bias across each TI layer, breaking the inversion symmetry between the two TI surfaces. According to Table 7.1, $\tilde{\Gamma} = \Gamma^\mu_4 \ (T \text{ even})$ or $\Gamma^\mu \ (T \text{ odd})$ where $\mu = 1, 2, 3, 5$. The local Green’s function has the decomposition,

$$G_{00}^0(\omega) = a(\omega)\mathbb{I} + b_1(\omega)\Gamma^4 + b_2(\omega)\tilde{\Gamma}.$$  \hspace{1cm} (7.33)

While this decomposition can be obtained analytically (see Eqs. D.14-D.15 and Eqs. D.17-D.18), its structure is easier to understand from the BCA: symmetry consideration demands that the local Hamiltonian $H_{00}^0 \equiv \frac{1}{N} \sum_k H_0^0(k) = \alpha \mathbb{I} + \beta \Gamma^4 + \eta \tilde{\Gamma}$ where the first two terms are the only possibilities to conserve both $T$ and $l$, see Table 7.1. Its inverse can potentially have four terms, $\mathbb{I}$, $\Gamma^4$, $\tilde{\Gamma}$, and $\Gamma^4\tilde{\Gamma}$. Since $\tilde{\Gamma}$ anticommutes with $\Gamma^4$, their cross term must vanish, yielding the form in Eq. 7.33. We will come back to BCA later.

The resonance condition can now be solved for different impurities. As an example, consider $\tilde{\Gamma} = \Gamma^\mu_4$ and $\Lambda = \Gamma^{\mu\bar{\mu}} = \Lambda^{-1}$ where $\bar{\mu} \in \{1, 2, 3, 5\} \setminus \{\mu\}$. This type commutes with $\Gamma^4$ but anticommutes with $\tilde{\Gamma}$, and includes the purely magnetic impurities $\Gamma_{12}, \Gamma_{23}$ and $\Gamma_{13}$. The Hermitian part of the inverse $T_{00}$ matrix
Figure 7.3: Stability of Weyl semimetal with \( I \) breaking \( \tilde{\Gamma} = \Gamma^{14} \) (bulk perturbation with both magnetic and orbital effects). The vertical axis \( \eta \) is the strength of the \( I \) breaking term. (a) stable zone for impurity classes \((+, -)\). (b) stable zone for impurity classes \((- , +)\). (c) stable zone for impurity class \((- , -)\). See Table 7.2 for impurity classification. Note how (c) resembles the union of (a) and (b), as could be predicted from Table 7.2. Solid black lines mark the two band edges bounding the central gap. They touch from around \( \eta = 0.88 \) to \( 1.35 \), corresponding to the Weyl semimetal phase. Dotted gray lines are the stable zone boundaries given by the band center approximation. It qualitatively agrees with the shape of the colored region near the central gap. The deviation is mainly deep in the bands (side wings in the colored region) where higher order terms in Eq. 7.31 become important. Parameters used are \( t = 0.05, t_1 = -0.5, t' = -0.25, \lambda = 3.5, \varepsilon_0 = -0.3 \), on a lattice of \( N_x = N_y = N_z = 50 \). Spectral broadening is \( \epsilon = 0.05 \).
$$T^{-1}_{00}(\omega) = g^{-1}\Gamma^{\mu\bar{\mu}} - a(\omega)\Gamma^4 - b_1(\omega)\Gamma^{\mu4}.$$  

(7.34)

Using the anticommutation \(\{g^{-1}\Gamma^{\mu\bar{\mu}} - b_1\Gamma^4, \Gamma^{\mu4}\} = 0\), the above can be rearranged into \((T^{-1}_{00} + a)^2 - b_2^2 = (g^{-1}\Gamma^{\mu\bar{\mu}} - b_1\Gamma^4)^2\). Setting \(T^{-1}_{00} = 0\), both sides can be simultaneously diagonalized, and the eigenvalues of the RHS are \((g^{-1} \pm b_1)^2\). The condition for vanishing \(T^{-1}_{00}\) is thus

$$g^{-1} = \pm b_1(\omega) \pm \sqrt{a(\omega)^2 - b_2(\omega)^2}.$$  

(7.35)

Resonance then requires \(g^{-1}\) to be real, \(viz.\ |a(\omega)| > |b_2(\omega)|\). The occurrence of a possibly negative term under the square root stems from the anticommutation of \(\bar{\Gamma}\) with \(\Lambda\), \(i.e.\) the interplay between the bulk symmetry breaking field and the impurity.

Similar analysis can be carried out when \(\Lambda\) is any of the sixteen \(\Gamma\) matrices. The results are summarized in the third column of Table 7.2. The sixteen \(\Gamma\)-matrix impurity candidates can be classified into four classes labeled by their commutation with \(\Gamma^4\) and \(\Gamma^{\mu4}\): \((+, -)\) denotes \(\Lambda\) commuting with \(\Gamma^4\) and anticommuting with \(\Gamma^{\mu4}\), and similarly for \((+, +), (-, +)\) and \((-,-)\). Impurities belonging to the same class have the same resonance condition. A nontrivial solution arises if there is at least one anticommutation, giving rise to a possibly negative term under the square root, and the protection of DOS suppression at Weyl nodes. The unperturbed \(H^0\) is parameterized by the symmetry-breaking strength \(\eta\), and one can ask how it affects the system’s ability to induce resonance at energy \(\omega\). The \((\omega, \eta)\) space is thus divided into two phases according to the existence of resonance. These are shown in Fig. 7.3, where the stable phases (no resonance) are colored.

The shape of the phase boundaries can be qualitatively understood in terms of the parameters of the Hamiltonian using BCA: the local Hamiltonian is \(H^0_{00} = \frac{1}{N} \sum_k H^0(k) = \alpha\mathbb{I} + \beta\Gamma^4 + \eta\Gamma^{\mu4}\) where \(\alpha = -\varepsilon_0, \beta = -12t' - \lambda\). Its Green’s function is \(\bar{G}(\omega) = \bar{a}(\omega)\mathbb{I} + \bar{b}_1(\omega)\Gamma^4 + \bar{b}_2(\omega)\Gamma^{\mu4}\) where \(\bar{a}(\omega) = (\omega - \alpha)/Q(\omega), \bar{b}_1(\omega) = \beta/Q(\omega), \bar{b}_2(\omega) = \eta/Q(\omega),\) and \(Q(\omega) = (\omega - \alpha)^2 - \beta^2 - \eta^2\). Note that these coefficients are independent of \(\bar{\Gamma}\), thus the stability of the Weyl nodes can be predicted according to the impurity class, regardless of \(\bar{\Gamma}\). The BCA version of the phase boundaries
Table 7.3: Impurity classification for $I$-symmetric (hence $T$ breaking) Weyl material. The class to which $\Lambda$ belongs are labeled by the two signs of the commutation of $\Lambda$ with $\Gamma^4$ and $\Gamma^{\mu\nu}$, respectively, where $+$ denotes commute and $-$ anticommute. The product of the two signs yields the commutation of $\Lambda$ with $\Gamma^4\Gamma^{\mu\nu}$. The two indices $\mu, \nu \in \{1, 2, 3, 5\}$ are fixed by the unperturbed Hamiltonian. The index $p \in \{1, 2, 3, 5\} \setminus \{\mu, \nu\}$. $s$ and $s'$ take the values of $\pm 1$. Solution of $g$ yielding $\det T - 1 = 0$ are summerized in the third column, see equations 7.37 and 7.41 in text. The resonance conditions for each class can be deduced by requiring $g^2$ to be positive (so that $g$ is real), and are explicitedly spelled out in the fourth column using BCA, which only need to be satisfied for either $s = 1$ or $-1$. Stability of the Weyl nodes are listed in the last column.

<table>
<thead>
<tr>
<th>$\Lambda$ class</th>
<th>$g^{-2}$</th>
<th>resonance (BCA)</th>
<th>stability</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mathbb{I}$, $\Gamma^4$, $\Gamma^{\mu\nu}$</td>
<td>$(a + sb_1 + s'b_2 + ss'b_3)^2$</td>
<td>any $\omega$</td>
<td>unstable</td>
</tr>
<tr>
<td>$\Gamma^{\mu p}$, $\Gamma^{\nu p}$</td>
<td>$-(a + sb_1)^2 - (b_2 + sb_3)^2$</td>
<td>$</td>
<td>\omega - \alpha - s\beta</td>
</tr>
<tr>
<td>$\Gamma^p$, $\Gamma^{4 p}$</td>
<td>$(a + sb_2)^2 - (b_1 + sb_3)^2$</td>
<td>$</td>
<td>\omega - \alpha - s\eta</td>
</tr>
<tr>
<td>$\Gamma^p$, $\Gamma^{4 p}$, $\Gamma^\nu$, $\Gamma^{4 \nu}$</td>
<td>$(a + sb_3)^2 - (b_1 + sb_2)^2$</td>
<td>$</td>
<td>\omega - \alpha</td>
</tr>
</tbody>
</table>

are shown as dotted lines in Fig. 7.3. The DOS suppression at the bulk Weyl nodes is protected for impurities in classes $(+, -)$, $(-, +)$ and $(-, -)$. The only unstable class is $(+, +)$, due to its fully-commuting nature with $G_{00}^0$.

7.5.3 $H^0(k)$ with $I$-symmetric $\tilde{\Gamma}$

To split the Dirac node into two Weyl nodes, an $I$-symmetric $\tilde{\Gamma}$ must break $T$. Thus $\tilde{\Gamma} = \Gamma^{\mu\nu}$ ($\mu \neq \nu \neq 4$) according to Table 7.1. The local Green’s function is

$$G_{00}^0(\omega) = a(\omega)\mathbb{I} + b_1(\omega)\Gamma^4 + b_2(\omega)\Gamma^{\mu\nu} + b_3(\omega)\Gamma^4\Gamma^{\mu\nu},$$

see Eqs. D.20-D.22 for expressions of the coefficients. The decomposition structure is easier to understand in terms of BCA: similar to the discussion beneath Eq. 7.33, one has $H_{00}^0 = a\mathbb{I} + \beta\Gamma^4 + \eta\tilde{\Gamma}$, thus its inverse has four possible terms, $\mathbb{I}$, $\Gamma^4$, $\tilde{\Gamma}$ and $\Gamma^4\tilde{\Gamma}$. Since $\Gamma^4$ and $\tilde{\Gamma}$ commute ($I$ symmetry), their cross term does not vanish, hence the form of Eq. 7.36.

Note that all three $\Gamma$ matrices in Eq. 7.36 mutually commute, and the product of any two is equal to the third. This implies that the impurity $\Lambda$ either commutes with all of them, or it commutes with one and anticommutes with the
other two (because the product of any two commutation signs should produce the third). In the fully commuting case, resonance can always be induced by impurities of strength

\[ g^{-1} = s_\Lambda \left[ a(\omega) + s_4 b_1(\omega) + s_{\mu\nu} b_2(\omega) + s_4 s_{\mu\nu} b_3(\omega) \right] \]  \hspace{1cm} (7.37)

where \( s_4, s_{\mu\nu} \) and \( s_\Lambda \) are eigenvalues of \( \Gamma^4, \Gamma^{\mu\nu} \) and \( \Lambda \) respectively and independently take the values \( \pm 1 \).

For other \( \Lambda \), there are two anticommutations. We can relabel the three \( \Gamma \) matrices in Eq. 7.36 according to their commutation with \( \Lambda \), and write the inverse \( T \) matrix as

\[ T^{-1}_{00} = -(aI + b_C \Gamma_C) + (g^{-1} - b_A \Gamma_A - b_{A'} \Gamma_{A'}) \]  \hspace{1cm} (7.38)

where \( [\Lambda, \Gamma_C] = \{ \Lambda, \Gamma_A \} = \{ \Lambda, \Gamma_{A'} \} = 0 \) and \( \{ \Gamma_C, \Gamma_A, \Gamma_{A'} \} \) is some permutation of \( \{ \Gamma^4, \Gamma^{\mu\nu}, \Gamma^4 \Gamma^{\mu\nu} \} \). The two parentheses in Eq. 7.38 mutually commute, thus \( T^{-1}_{00} \) is block-diagonal: in the eigen-subspace of \( \Gamma_C \) with eigenvalue \( \pm 1 \), the matrices \( \Gamma_C, \Lambda, \Gamma_A \) and \( \Gamma_{A'} \) reduce to \( 2 \times 2 \) blocks, denoted as \( \pm I, \Lambda_{\pm}, \Gamma_{A\pm} \) and \( \Gamma_{A'\pm} \), respectively, all of which square to \( I \). Since the projectors onto the subspaces of \( \Gamma_C \) commute with \( \Gamma_A, \Gamma_{A'} \) and \( \Lambda \), the mutual (anti)commutation relations of the latter four are inherited in both subspaces. Setting \( T^{-1}_{00} = 0 \) in Eq. 7.38 for both blocks then yields

\[ (a \pm b_C)I = g^{-1} - b_A \Gamma_A - b_{A'} \Gamma_{A'} \]  \hspace{1cm} (7.39)

Squaring both sides and using the fact that \( \Gamma_{A\pm} \Gamma_{A'\pm} = \pm I \), which follows from \( \Gamma_A \Gamma_{A'} = \Gamma_C \), one gets

\[ g^{-2} = (a \pm b_C)^2 - (b_A \pm b_{A'})^2 \]  \hspace{1cm} (7.40)

The resonance condition is for \( g \) to be real, \textit{viz.},

\[ |a \pm b_C| > |b_A \pm b_{A'}| \]  \hspace{1cm} (7.41)

here the two \( \pm \) signs are correlated, and the reality of \( g \) requires only one of the two branches to be satisfied. This is enumerated in the third column in Table 7.3 and plotted in Fig. 7.4.
Figure 7.4: Stability of Weyl semimetal with $I$-symmetric and $T$ breaking $\tilde{\Gamma} = \Gamma^{12}$ (external magnetic field). The vertical axis $\eta$ is the strength of the $T$ breaking term. Green (light shade): stable zone for impurity classes $(+, -)$ and $(-, -)$. Red (dark shade): stable zone for impurity classes $(-, +)$ and $(-, -)$. See Table 7.3 for impurity classification. Solid black lines mark the two band edges bounding the central gap. They touch from around $\eta = 1$ to 5, corresponding to the Weyl semimetal phase. The black lines are broken around $\eta = 1$ due to the closing of the indirect gap. Dotted gray lines are the stable zone boundaries given by the band center approximation. Parameters used are $t = 0.05, t_1 = -0.5, t' = -0.25, \lambda = 1, \epsilon_0 = -0.3$, on a lattice of $N_x = N_y = N_z = 50$. Spectral broadening is $\epsilon = 0.05$. 
As before, the expansion coefficients in Eq. 7.36 can be estimated by the BCA and used to approximate the boundaries between the resonant and non-resonant phases in the \((\omega, \eta)\) space. The local Hamiltonian is \(H_{00} = \alpha \mathbb{I} + \beta \Gamma^4 + \eta \Gamma^{\mu \nu}\) where \(\alpha = -\varepsilon_0\), \(\beta = -12t' - \lambda\). Its Green’s function is \(\bar{G}(\omega) = \bar{a}(\omega) \mathbb{I} + \bar{b}_1(\omega) \Gamma^4 + \bar{b}_2(\omega) \Gamma^{\mu \nu} + \bar{b}_3(\omega) \Gamma^4 \Gamma^{\mu \nu}\) with

\[
\bar{a}(\omega) = (\omega - \alpha) \left[ (\omega - \alpha)^2 - \beta^2 - \eta^2 \right] / Q(\omega) \\
\bar{b}_1(\omega) = \beta \left[ (\omega - \alpha)^2 - \beta^2 + \eta^2 \right] / Q(\omega) \\
\bar{b}_2(\omega) = \eta \left[ (\omega - \alpha)^2 + \beta^2 - \eta^2 \right] / Q(\omega) \\
\bar{b}_3(\omega) = 2(\omega - \alpha) \beta \eta / Q(\omega)
\] (7.42)

where \(Q(\omega) = [(\omega - \alpha)^2 - (\beta + \eta)^2][(\omega - \alpha)^2 - (\beta - \eta)^2]\). The resulting resonance conditions are summarized in the fourth column of Table 7.3, and the conditions for stable Weyl nodes (if exist) in the last column.

The stable zones of the two classes \((+, -)\) (green in Fig. 7.4) and \((-+, +)\) (red in Fig. 7.4) are restricted to opposite sides of a critical value of the symmetry-breaking strength \(\eta = \eta_c\). Furthermore, near \(\eta_c\), the region of stable energy narrows down toward the Weyl node. One can think of the resonance energy as forming an impurity band generated by a continuum of impurity strengths \(g\). Then the stable zones constitute gaps in such bands. In this sense, \(\eta_c\) marks a phase transition of the impurity band from gapless to gapped. The existence of \(\eta_c\) can be understood from the BCA, according to which the phase boundaries are given by

\[
\omega = -\varepsilon_0 \pm \left( |\beta| - |\eta| \right),
\] (7.43)

shown as gray dotted lines in Fig. 7.4. These are the two central band centers (eigenvalues of \(H_{00}^0\)). They cross at \(\eta = |\beta|\), which gives the critical strength \(\eta_c\). \(\eta_c = 2\) in Fig. 7.4. This is reminiscent of the bulk band inversion in topological/Chern insulators that signifies a gapless to gapped transition in their surface/edge spectrum.

To illustrate the above impurity band phase transition, we employ the average \(T\)-matrix approximation (ATA) to investigate the effect of an ensemble of local impurities spatially uniformly distributed with concentration \(c\) [19, 3]. The
Figure 7.5: Effect of different classes of impurity ensembles on the ATA DOS. The color map shows $\log |T_{00}(\omega + i\epsilon)||$, where darker color denotes larger $|T_{00}||$. Trails of darkest color will follow the zeros of $T^{-1}_{00}$ as given by the third column in Table 7.3. Black dash-dot curves are DOS of the clean system (identical in all four panels). Blue solid curves are DOS after adding impurity ensembles, and are computed using the average $T$-matrix approximation (ATA). Impurity class used in each panel are given in their respective caption. Impurity strengths are uniformly distributed in $g \in (0, 10]$. Impurity concentration is $c = 10\%$. The dotted horizontal line marks the minimum of $g^{-1}(0.1)$: the ATA DOS is significantly enhanced for those $\omega$ where the high $\log |T_{00}||$ lines exist above this line. $\eta = 3$ is used; according to Fig. 7.4, the clean system is in the WS phase, and the DOS at the Weyl node should be suppressed for classes $(+, -)$ (b) and $(-, -)$ (d) but enhanced for the other two classes. This agrees with the plots shown here. Other parameters are the same as in Fig. 7.4: $t = 0.05, t_1 = -0.5, t' = -0.25, \lambda = 1, \varepsilon_0 = -0.3$, on a lattice of $N_x = N_y = N_z = 50$. The spectral broadening (imaginary part of $\omega$) is taken to be $\epsilon = 0.05$, which prevents the DOS from touching zero (as it should) at the Weyl node energy near $\omega = 0.4$. 
entire ensemble has the same matrix form, but with strength \( g \) given by some distribution \( f(g) \). In the ATA formalism, statistical averaging over the \( f(g) \) will restore translational symmetry. The impurity effect is then captured by a local self energy, \( \Sigma_{\text{loc}} = c(T_{00})[1+cG_{00}^0(T_{00})]^{-1} \), where \( \langle \cdots \rangle \) denotes the \( f(g) \) averaging. The self-energy-corrected local Green’s function is 
\[
G_{\text{loc}}(z) = \frac{1}{N} \sum_k \frac{1}{z - H^0(k) - \Sigma_{\text{loc}}}
\]
and the average LDOS is 
\[
\rho(\omega) = -\text{Im} \text{Tr} G_{\text{loc}}(\omega + i\eta^+).\]
One then expect the ATA DOS, \( \rho_{\text{ATA}} \), to be enhanced from the clean fraction, \( (1-c)\rho_{\text{clean}} \), for \( \omega \) in the unstable phase but reduced in the stable phase (since the integrated DOS is conserved). This is shown in Fig. 7.5, in which we plot at a fixed \( \eta \) the simplest case where \( f(g) \) is a constant for \( g \in (0,10] \) and zero otherwise. For this particular \( \eta \) value, the Weyl nodal energy is stable in (b) and (d), but is unstable in (a) and (c).

7.6 Summary and discussion

In this chapter, we study the effect of localized impurities \( V = g\Lambda\delta(\mathbf{x}) \) on the bulk electronic structure of Weyl semimetals. A general method is devised to detect whether or not a resonance can be induced at energy \( \omega \) by a \( \Lambda \)-type impurity. If such a resonance is possible, then \( \omega \) is said to be unstable with respect to \( \Lambda \), otherwise it is stable. The stability of \( \omega \) requires all eigenvalues of \( G_{00}^0(\omega)\Lambda \) to have finite imaginary part. Here, \( G_{00}^0(\omega) \) is the Hermitian part of the retarded local Green’s function \( G_{00}(\omega + i\eta^+) \). Otherwise, one can always use a coupling strength \( g = 1/u_a(\omega) \) to induce a resonance at \( \omega \), with \( u_a(\omega) \) being the purely real eigenvalue of \( G_{00}^0(\omega)\Lambda \) indexed by \( a \). The existence of real \( u_a(\omega) \) is equivalent to requiring the Hermitian part of the inverse \( T \) matrix to have a zero eigenvalue. An immediate corollary is that impurities commuting with \( G_{00}^0 \) can induce a resonance at an arbitrary energy, simply by tuning the impurity strength. This includes the physically important class of local chemical potential perturbations.

We applied this method to four-band lattice Weyl semimetal models, expressed in terms of Dirac \( \Gamma \) matrices. For these models, the \( T \) matrix and its eigenvalues can be obtained analytically. Mathematically, one first classifies the clean Weyl semimetals according to whether or not inversion (I) is broken. The
difference is in the decomposition of their local Green’s functions $G_{00}^0$. In each case, impurities are then classified by their commutations with the $\Gamma$ matrices appearing in $G_{00}^0$: anticommutation with components of $G_{00}^0$ result in a square-root structure, which constrains the reality of $u_a(\omega)$. Note that in this scheme, it is more relevant to know which $\Gamma$ matrices appear than their exact numerical coefficients. For this purpose the band center approximation (BCA) – which replaces $G_{00}^0(\omega)$ by $(\omega - H_{00}^0)^{-1}$ where $H_{00}^0$ is the local Hamiltonian – is quite useful as it has the same form of $\Gamma$ matrix decomposition with coefficients whose meanings are physically more transparent. Results for $I$ breaking WS are reported in Table 7.2, and for $I$ invariant WS in Table 7.3.

Realistic impurities are more likely to be linear combinations of multiple $\Gamma$ matrices, mixing orbital, magnetic, and chemical potential effects. A linear combination of impurities in the same class, if it still squares to identity, is no different from a single $\Gamma$ matrix in that class, and results obtained before hold unchanged. While other combinations are not studied here, it is reasonable to expect that stability will resemble that of the dominant component if there is one, and crossover will happen as the relative strengths change. We have confirmed this for several tractable cases of Dirac semimetals. The method for obtaining the relation between impurity strength $g$ and the induced resonance/bound state energy $\omega$ may also prove useful in device engineering where specific energy levels are desired. For Dirac materials with random strength disorder, results similar to those shown in Fig. 7.5 are expected, where roughly speaking impurity induced states form their own band superimposed on the clean DOS, and stable energies constitute the band gap. Such impurity bands may modify transport properties if certain impurity “superlattice” is approximately formed, or if the coherent length of single-impurity resonances become compatible with the impurity density.

Chapter 8

Generalized Kitaev models on three dimensional lattices

In a seminal paper [38], Kitaev proposed a model of interacting spin-$1/2$s on a honeycomb lattice, which can be solved exactly by a mapping of the spins to free Majoranas hopping in static $Z_2$ gauge fields. To find the ground state, for example, one diagonalizes the quadratic Majorana Hamiltonian and finds the lowest energy state for each $Z_2$ configuration. The true ground state is then the global minimum of these lowest energy states. In fact the energy of any desired flux sector could be lowered arbitrarily by adding certain types of interaction terms to the Hamiltonian which do not affect solvability, thus being a true ground state is no longer a very restrictive requirement. Kitaev’s solution relies on a match between the honeycomb lattice coordination number and the rank of the Clifford algebra underlying the spin-$1/2$ Pauli matrices, thus in its generalization to other lattices, it may become necessary to replace the Pauli matrices with higher rank generalized $\Gamma$ matrices, e.g., rank-4 Dirac matrices for lattices of coordination number 5. The advent of Kitaev’s model and its various generalizations has led to an embarrassment of riches: suddenly topological phases and spin metals become easy to generate.
8.1 Kitaev’s Honeycomb model

In [38], Kitaev introduced an exactly solvable model of interacting spin-$\frac{1}{2}$s on a honeycomb lattice with Hamiltonian

$$H = -\sum_{\langle ij \rangle \in x} J_x \sigma^x_i \sigma^x_j - \sum_{\langle ij \rangle \in y} J_y \sigma^y_i \sigma^y_j - \sum_{\langle ij \rangle \in z} J_z \sigma^z_i \sigma^z_j.$$  \hspace{1cm} (8.1)

Here, $x, y, z$ denotes the three directions of nearest neighbor bonds on the honeycomb lattice (Fig. 8.1) \(^1\), $i, j$ are site indices. $\sigma^{x,y,z}_i$ are the Pauli matrices defined on each site, and should be understood as shorthands for direct products,

$$\sigma^{x,y,z}_i = \underbrace{I \otimes I \otimes \cdots \otimes I \otimes \sigma^{x,y,z}}_{i-1} \otimes \cdots.$$ \hspace{1cm} (8.2)

For each hexagonal plaquette, one can define

$$W \equiv \sigma^x_1 \sigma^z_2 \cdot \sigma^y_2 \sigma^z_3 \cdot \sigma^z_3 \sigma^z_4 \cdot \sigma^x_4 \sigma^y_5 \cdot \sigma^y_5 \sigma^z_6 \cdot \sigma^z_6 \sigma^x_1 = -\sigma^y_1 \sigma^z_2 \sigma^z_3 \sigma^y_4 \sigma^z_5 \sigma^x_6.$$ \hspace{1cm} (8.3)

\(^1\)While the designation of $x, y, z$ seems contrived, note that a cubic lattice does flatten into a honeycomb when looked at along the $(1,1,1)$ direction.

Figure 8.1: Kitaev’s honeycomb model. $x, y,$ and $z$ links are colored in red, green, and blue, respectively. 1 to 6 label the vertexes of a hexagonal plaquette (Eq. 8.3).
where 1 · · · 6 label the sites, with the link between 1 and 2 along the \(x\) direction, etc. (Fig. 8.1). One can verify that \(W\) on different plaquettes commute, and they also commute with \(H\). For a lattice of \(N\) unit cells, there are \(2N\) sites and hence \(2^{2N}\) degrees of freedom (DOF). Since there are as many plaquettes as there are unit cells and each plaquette accounts for a DOF of 2 (which follows from \(W^2 = 1 \implies W = \pm 1\)), the \(N\) plaquettes account for \(2^N\) DOFs. As we shall see, the remaining \(2^N\) DOFs can be mapped to free fermions hopping in a background of fixed \(\{W\}\). The Kitaev problem is thus exactly solvable.

8.1.1 Fermionization

Consider the following fermionization,

\[
\tilde{\sigma}_i^a \equiv i\theta_0^a \theta_i^a , \quad a = 1, 2, 3 \equiv x, y, z .
\]  

(8.4)

where \(\theta_i^\mu (\mu = 0, 1, 2, 3)\) are Majoranas defined on site \(i\) with anticommutation

\[
\{\theta_i^\mu , \theta_j^\nu \} = 2\delta_{ij}\delta^{\mu\nu}.
\]  

(8.5)

The \(\tilde{\sigma}\)'s thus have the same commutation relations as the \(\sigma\)'s. However, they do not preserve the Pauli algebra: \(\sigma^x \sigma^y \sigma^z = i\), whereas \(\tilde{\sigma}^x \tilde{\sigma}^y \tilde{\sigma}^z = iD\) with

\[
D_i = D_i^\dagger \equiv \theta_0^1 \theta_i^1 \theta_0^2 \theta_i^2 .
\]  

(8.6)

One can verify that

\[
[D_i , D_j] = [D_i , \tilde{\sigma}_j] = 0 \quad \forall i, j .
\]  

(8.7)

By Eq. 8.5, \(D_i^2 = 1\), thus \(D_i\) has eigenvalues \(\pm 1\), and Pauli algebra is preserved in the \(D_i = 1\) subspace. Let \(P_i\) project onto \(D_i = +1\),

\[
P_i = (1 + D_i)/2 ,
\]  

(8.8)

then one can make the following identification,

\[
\sigma_i^a = P_i \tilde{\sigma}_i^a P_i = P_i \tilde{\sigma}_i^a ,
\]  

(8.9)
where the second equality follows from Eq. 8.7. This implies
\[ H = P \tilde{H} P = P \tilde{H} \] (8.10)
where \( P = \prod_i P_i \), \( \tilde{H} \) is obtained by replacing the \( \sigma \)'s with \( \tilde{\sigma} \)'s in \( H \),
\[ \tilde{H} = -\sum_{\langle ij \rangle \in x} J_x \tilde{\sigma}_i^x \tilde{\sigma}_j^x - \sum_{\langle ij \rangle \in y} J_y \tilde{\sigma}_i^y \tilde{\sigma}_j^y - \sum_{\langle ij \rangle \in z} J_z \tilde{\sigma}_i^z \tilde{\sigma}_j^z , \] (8.11)
and we have used \([\tilde{H}, P] = 0\) (which follows from Eq. 8.7) to obtain the second equality of Eq. 8.10. Thus if \(|\tilde{\Psi}\rangle\) is an eigenstate of \( \tilde{H} \), then \( P|\tilde{\Psi}\rangle\) is an eigenstate of \( H \) with the same energy. Note that degenerate eigenstates of \( \tilde{H} \) may correspond to the same eigenstate of \( H \) (e.g. gauge redundancy). The task now is to solve the eigenvalue problem of \( \tilde{H} \).

### 8.1.2 Fermionization, first quantized form

It is instructive to translate the above second-quantized language back to matrices in the tensor product space in which the original Hamiltonian Eq. 8.1 is formulated. On each site, introduce two Dirac fermions \( c \) and \( d \),
\[ c = (\theta^0 + i\theta^1)/2 \quad , \quad d = (\theta^2 + i\theta^3)/2 , \] (8.12)
which satisfy \( \{c, c^\dagger\} = \{d, d^\dagger\} = 1 \). Then
\[ \theta^0 = c + c^\dagger \quad , \quad \theta^1 = (c - c^\dagger)/i \quad , \quad \theta^2 = d + d^\dagger \quad , \quad \theta^3 = (d - d^\dagger)/i . \] (8.13)
The complete basis on each site is \(|\emptyset\rangle, |c\rangle, |d\rangle, |cd\rangle\) where \(|cd\rangle = c^\dagger d^\dagger |\emptyset\rangle\), etc., and \(|\emptyset\rangle\) is the vacuum. In this basis,
\[ c = \begin{pmatrix} \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & 1 \end{pmatrix} = \frac{1}{2} \mathbb{I} \otimes \tau^+ \quad , \quad d = \begin{pmatrix} \cdot & 1 & \cdot \\ \cdot & \cdot & -1 \\ \cdot & \cdot & \cdot \end{pmatrix} = \frac{1}{2} \tau^+ \otimes \tau^z \] (8.14)
where \( \cdot \cdot \cdot \) denotes zero, and where \( \tau \) are Pauli matrices, with \( \tau^\pm = \tau^x \pm i\tau^y \).
Eq. 8.13 then gives
\[ \theta^0 = \mathbb{I} \otimes \tau^x \quad , \quad \theta^1 = \mathbb{I} \otimes \tau^y \quad , \quad \theta^2 = \tau^x \otimes \tau^z \quad , \quad \theta^3 = \tau^y \otimes \tau^z , \] (8.15)
thus
\[ \tilde{\sigma}^x = -\mathbb{I} \otimes \tau^x \quad , \quad \tilde{\sigma}^y = \tau^x \otimes \tau^y \quad , \quad \tilde{\sigma}^z = \tau^y \otimes \tau^y \]  
(8.16)
and
\[ D = -\tau^z \otimes \tau^z = \text{diag}(-1, 1, 1, -1) \]  
(8.17)
The $D = 1$ subspace corresponds to the one spanned by $\{|c\rangle, |d\rangle\}$, i.e. the one with a single Dirac fermion per site. Projected onto this space one has
\[ P\tilde{\sigma}^x P = \tau^z \quad , \quad P\tilde{\sigma}^y P = -\tau^y \quad , \quad P\tilde{\sigma}^z P = \tau^x . \]  
(8.18)
Thus the interpretation that $\sigma = P\tilde{\sigma} P$ implies that the eigenstates of $\tau^z$ above are $|c\rangle = |\rightarrow\rangle$ and $|d\rangle = |\leftarrow\rangle$, i.e. fermion states with spins pointing along $x$ and $-x$.

### 8.1.3 $Z_2$ block-diagonalization

In terms of the Majoranas, the Hamiltonian in the enlarged space is
\[ \tilde{H} = i \sum_{\langle ij \rangle \in x} J_x \theta_i^0 \theta_j^0 \hat{u}_{ij} + i \sum_{\langle ij \rangle \in y} J_y \theta_i^0 \theta_j^0 \hat{u}_{ij} + i \sum_{\langle ij \rangle \in z} J_z \theta_i^0 \theta_j^0 \hat{u}_{ij} , \]  
(8.19)
where the operators $\hat{u}_{ij}$ live on the honeycomb links,
\[ \hat{u}_{ij} \equiv i \theta_i^a \theta_j^a \quad \text{for} \quad \langle ij \rangle \in a \quad , \quad a = x, y, z . \]  
(8.20)
The set of $\{\hat{u}_{ij}\}$ on all links mutually commute, and they also commute with $\tilde{H}$. Thus $\tilde{H}$ can be block-diagonalized into sectors labeled by the eigenvalues $\{u_{ij}\}$ of the link operators, which are $\pm 1$ following $\hat{u}_{ij}^2 = 1$. Within each sector, the Hamiltonian reduces to a tight-binding Hamiltonian of the $\{\theta^0\}$ Majoranas hopping in a static $\{u_{ij}\}$ configuration,
\[ \tilde{H}_{\{u\}} = i \sum_{\langle ij \rangle} A_{ij} \theta_i \theta_j \quad , \quad A_{ij} = J_{ij} u_{ij} , \]  
(8.21)
where $J_{ij} = J_{x,y,z}$ depending on the bond direction, and we have dropped the flavor index “0” from the $\theta$’s. Note that the matrix $A$ is real and antisymmetric.
There are some peculiarities of such bilinear Majorana Hamiltonians, which we shall discuss later in §8.2.

The operator $D_i$ acts as a local gauge transformation on site $i$ by flipping the sign of all link operators emanating from site $i$,

$$D_i^\dagger \hat{u}_{ij} D_i = -\hat{u}_{ij}, \quad (8.22)$$

while leaving both $\tilde{H}$ and $\tilde{H}_{\{u\}}$ invariant (because $D_i \theta_i D_i = -\theta_i$ provides another minus sign). Thus sectors labeled by $\{u\}$ and $D_i \{u\} D_i$ are physically equivalent.

The gauge invariant content is the set of plaquette operators $\{W\}$ (Eq. 8.3). Using $\sigma_a^i \equiv \tilde{\sigma}_a^i = \tilde{\sigma}_a^i P$, one has

$$\hat{W} \equiv \hat{u}_{12} \hat{u}_{23} \hat{u}_{34} \hat{u}_{45} \hat{u}_{56} \hat{u}_{61}, \quad W = P \hat{W} = \hat{W} P. \quad (8.23)$$

$W$ is invariant under the action of any $D_i$ since it only affects two bonds on a plaquette containing site $i$, and leaves the above product unchanged. The projection $P$ does not commute with the link operators (Eq. 8.22), but it commutes with the plaquettes $W$ (Eq. 8.23). In other words, the gauge fluxes $W$ are physical, but the gauge fields $\hat{u}$ are not.

Finally, physical eigenstates are obtained by applying $P$ to eigenstates of $\tilde{H}_{\{u\}}$. If $|\tilde{\Psi}_u\rangle$ is an eigenstate of $\tilde{H}_{\{u\}}$, then the corresponding eigenstate of $H$ should be labeled by the gauge flux $W$ realized by the configuration $u$, and is, up to an overall normalization,

$$|\Psi_W\rangle = P|\Psi_u\rangle \propto \prod_i (1 + D_i)|\Psi_u\rangle = \sum_{u' \sim u} |\Psi_{u'}\rangle, \quad (8.24)$$

where $u' \sim u$ denotes all gauge configurations equivalent to $u$, and can be any configuration obtained by applying $D_i$ on an arbitrary subset of sites. The last equality is obtained by expanding $\prod_i (1 + D_i)$, and noting that $D_i |\Psi_u\rangle = |\Psi_{D_i u D_i}\rangle$. Thus the physical state of a given flux configuration is an equal-weight superposition of the corresponding unprojected states of all possible gauge field realizations.

In the rest of this chapter, we will focus on the fermion aspect of the Kitaev model and its generalizations.
8.2 Free Majorana Hamiltonians

We briefly summerize the diagonalization of quadratic Majorana Hamiltonians in this section.

8.2.1 Generic case

Let \( \{ \theta_i \} \) be a set of \( N \) Majorana fermions,
\[
\theta_i = \theta_i^\dagger, \quad \{ \theta_i, \theta_j \} = 2\delta_{ij}, \quad i, j = 1, 2, \cdots, N.
\] (8.25)

A quadratic Majorana Hamiltonian can be written as
\[
\hat{H} = H_{ij} \theta_i^\dagger \theta_j, \quad H \equiv iA, \quad A^* = -A^T.
\] (8.26)

Since \( H^* = -H \), its spectrum is particle-hole symmetric. For even \( N \equiv 2M \), \( H \) is diagonalized by
\[
U = (|u_1\rangle, |u_2\rangle, \cdots |u_M\rangle; |u_1\rangle^*, |u_2\rangle^*, \cdots |u_M\rangle^*) ,
\] (8.27)

\[
U^\dagger H U = \begin{pmatrix} E & \cdots \\ \cdots & -E \end{pmatrix}, \quad E = \text{diag}(\varepsilon_1, \varepsilon_2, \cdots \varepsilon_M), \quad \varepsilon_n \geq 0,
\] (8.28)

where \( |u_n\rangle \) are column vectors. The diagonal basis of \( \hat{H} \) are the \( \{ \phi \} \) fermions,
\[
\phi_n = (U^\dagger)_{ni} \theta_i , \quad \phi^{\dagger}_{m+N} = \phi^*_m = \phi^*_m \quad (n = 1, \cdots, N = 2M \text{ and } m = 1, \cdots, M),
\] (8.29)

\[
\{ \phi^{\dagger}_n, \phi_n \} = 2\delta_{m,n} \implies \{ \phi_m, \phi_n \} = 2\delta_{m,n} \quad (m, n = 1, 2, \cdots, N = 2M),
\] (8.30)

where we have used the notation \( \phi_n \equiv \phi^*_n \). The independent subset of \( \{ \phi \} \) form \( M \) Dirac fermions,
\[
c_n \equiv \frac{\phi_n}{\sqrt{2}}, \quad \{ c^\dagger_m, c_n \} = \delta_{m,n} , \quad \{ c_m, c_n \} = 0 \quad (m, n = 1, 2, \cdots, M).
\] (8.31)

The Hamiltonian in the diagonal basis is
\[
\hat{H} = \sum_{n=1}^{M} \varepsilon_n (\phi^\dagger_n \phi_n - \phi^\dagger_n \phi_n) = 2 \sum_{n=1}^{M} \varepsilon_n (\phi^\dagger_n \phi_n - 1) \quad (8.32)
\]
\[
= 4 \sum_{n=1}^{M} \varepsilon_n \left( \hat{N}_n - \frac{1}{2} \right), \quad \hat{N}_n = c^\dagger_n c_n.
\] (8.33)
The ground state is the $c$-vacuum with all $N_n = 0$ (note that $\varepsilon_n \geq 0$).

For odd-dimensional $H$ with $N = 2M + 1$, particle-hole symmetry demands the extra eigenvalue to be zero with real eigenvector $|u_0\rangle$, and the diagonalized Hamiltonian can formally be written as

$$\hat{H} = \hat{H}_{2M} + 0\phi_0\phi_0 \quad , \quad \phi_0 = (U^\dagger)_{0i}\theta_i = \sum_i \langle i|u_0\rangle^\ast \theta_i = \phi_0^\dagger \quad (8.34)$$

where $\hat{H}_{2M}$ denotes the non-zero spectral part, and $\phi_0 = \phi_0^\dagger$ is the unpaired Majorana representing a $\sqrt{2}$-degree of freedom. Two such modes may pair into a Dirac fermion and generate a $\sqrt{2}^2$-fold degeneracy for all many-body states (including the ground state).

Note that all zero modes are of Majorana nature regardless of the parity of $N$. If $\varepsilon_n = 0$, then $|u_n\rangle$ and $|u_n\rangle^\ast$ are either equivalent or linearly independent. In the former case, $\phi_n = \phi_n^\dagger$ is a Majorana. In the latter case, $(\phi_n + \phi_n^\dagger)/\sqrt{2}$ and $(\phi_n - \phi_n^\dagger)/\sqrt{2}i$ are two independent Majoranas.

### 8.2.2 Translationally invariant case

On a lattice with a basis, the Majoranas can be indexed as $\theta^a_r$ where $r$ and $a$ denote unit cell and internal space coordinate, respectively. Let $\theta_r = \{\theta^a_r\}$, one has

$$\hat{H} = iA_{rr}\theta_r\theta^\dagger_r \quad , \quad A_{rr} = A^\ast_{rr} = -A^t_{rr} \quad , (8.35)$$

where $t$ denotes internal space transposition. With translational invariance, we have

$$\hat{H} = \sum_k \hat{H}_k \quad , \quad \hat{H}_k = iA_k\theta_{-k}\theta^\dagger_k \quad , \quad A_k = \sum_r A_0 e^{ikr} \quad , (8.36)$$

$$\implies A_k = -A^t_{-k} = -A^\dagger_k = A^\ast_{-k} \quad , (8.37)$$

$$\theta_r = \frac{1}{\sqrt{N}} \sum_k e^{ikr}\theta^\dagger_k \quad , \quad \theta^\dagger_k = \frac{1}{\sqrt{N}} \sum_r e^{-ikr}\theta_k = \theta_{-k} \quad , (8.38)$$

$$\implies \{\theta^a_k, \theta^b_{k'}\} = 2\delta_{k+k',0} \delta_{a,b} \quad . (8.39)$$

\[2\text{Thus Majorana fermions are said to have a quantum dimension of } \sqrt{2}. \text{ In general, the Hilbert space of } N \text{ particles with quantum dimension } q \text{ is } q^N \text{-dimensional.}\]

\[3\theta \text{ transforms like } c \text{ instead of } c^\dagger.\]
Note that at inversion-invariant momenta $Q$, $\hat{H}_Q$ is a free Majorana Hamiltonian (viz., $\theta_Q^\dagger = \theta_Q$ and $A_Q = A_Q^* = -A_Q^t$). $\hat{H}_k$ can be diagonalized as

$$\hat{H}_k = \sum_n \varepsilon_{k,n} \phi_{k,n}^\dagger \phi_{k,n}, \quad \text{diag}(\varepsilon_{k,1}, \varepsilon_{k,2}, \ldots) = U_k^\dagger (iA_k) U_k, \quad (8.40)$$

$$\phi_k = U_k^\dagger \theta_k \implies \{ \phi_{k,m}, \phi_{k',n} \} = 2 \delta_{k,k'} \delta_{m,n} \quad (8.41)$$

and $U_k$ and $U_{-k}$ are complex conjugates for $k \notin \{Q\}$, which follows from $iA_k = -(iA_{-k})^*$,

$$U_{-k} = U_k^* \implies \phi_{-k} = \phi_k^\dagger, \quad \varepsilon_{-k,n} = -\varepsilon_{k,n}, \quad (k \notin \{Q\}) \quad (8.42)$$

i.e., the collective index $(k, n)$ is paired with $(-k, n)$. Note that eigenvalues at $k$ and $-k$ are organized in opposite orders.

For $k \in \{Q\}$, the discussion of §8.2.1 applies. The diagonal basis of $\hat{H}_Q$ is $\{\phi_{Q,n}, \phi_{Q,n} = \phi_{Q,n}^*\}$ with corresponding energies $\{\varepsilon_{Q,n}, \varepsilon_{Q,n} = -\varepsilon_{Q,n}\}$ for even-membered unit cells, and an extra zero Majorana mode $\phi_{Q,0} = \phi_{Q,0}^*$ for odd-membered unit cells. For $n \neq 0$, $\phi_{Q,n}$ is paired with $\phi_{Q,n}$. For the zero-energy Majoranas, if $N_Q$ (the number of $Q$ points) is even, then $\phi_{Q,0}$ at different $Q$ can pair into zero-energy Dirac fermions, e.g., $Q = 0$ can be paired with $Q = \pi$ for a 1D system—in higher dimensions, there are $C_{N_Q}^2$ pairing schemes in general. If $N_Q$ is odd, then one $\phi_{Q,0}$ is unpaired. This could arise in Kitaev models with odd number of lattice sites, the implication of which is unclear.

For reference, we write down the full Hamiltonian in the diagonal basis where the number of basis elements is odd, $N_B = 2M_B + 1$,

$$\hat{H} = \sum_k' \sum_{n=1}^{N_B} 2\varepsilon_{k,n}(2c_{k,n}^\dagger c_{k,n} - 1) + \sum_Q \sum_{n=1}^{M_B} 2\varepsilon_{Q,n}(2c_{Q,n}^\dagger c_{Q,n} - 1) + 0 \cdot \sum_Q \phi_{Q,0} \phi_{Q,0} \quad (8.43)$$

where $\sum_k'$ denotes summation over half of the non-inversion-invariant $k$ points (one from each $\pm k$ pair). At $k \in Q$, only half of the $n$ indices are summed. The last term serves as a reminder that there are $N_Q$ Majorana eigen-modes guaranteed by $N_B$ being odd, and represent a $\sqrt{2}^{N_Q}$-fold degeneracy. If $N_Q$ is odd, it is anomalous as discussed before. If $N_B$ is even, then it will “vanish”. 

Figure 8.2: 1D Spin lattice. (a) Arrows on bonds denote positive direction. Horizontal bonds of $A$ and $B$ sublattices are of strengths $J$ and $sJ$ respectively with $s = \pm 1$, and vertical bonds have strength 1. Flux per plaquette is $\arg(i^4 s) = (1-s)\pi/2$. Ground state must have $\pi$ flux ($s = -1$) by Lieb’s theorem. (b) Double-periodicity modulation in horizontal bonds for the 0-flux case.

8.3 1D model: spin ladder

In this section, we use a 1D toy model to clarify some ambiguities inherent in models of Majorana fermions.

Consider a 1D spin lattice shown in Figure 8.2(a). The $k$-space Hamiltonian is

$$H(k, s) = \sigma_y - 2J \sin k \begin{pmatrix} 1 \\ s \end{pmatrix},$$

(8.44)

where $s = \pm 1$ denotes the direction of the $B$ sublattice bonds. The energy spectrum is

$$E(k) = \begin{cases} -2J \sin k \pm 1, & s = +1 \text{ (0 flux per square plaquette)} \\ \pm \sqrt{1 + 4J^2 \sin^2 k}, & s = -1 \text{ (}\pi\text{ flux per square plaquette)} \end{cases}.$$

(8.45)

8.3.1 $\pi$-flux sector is generically gapped

According to Lieb’s Theorem [42], the $\pi$-flux sector is the ground state. In this particular model, it is fully gapped and has no Fermi surface (points in 1D). However, even if a Fermi surface is present, it will be unstable in the following sense. We first note that the $\pi$-flux sector is inversion symmetric,

$$\sigma_y H_k \sigma_y = H_{-k}.$$

(8.46)
By inversion symmetric, we mean that the system is \emph{gauge equivalent} before and after lattice inversion. Here, lattice inversion is $\sigma_x$, and the required gauge transformation is $\sigma_z$, hence the combined unitary transformation $\sigma_y$. Since Majorana Hamiltonians satisfy $H_k = -H_k^*$ generically (Eq. 8.36), we have that $\text{spec}(H_k) = -\text{spec}(H_k)$, i.e., there is a spectral symmetry at each $k$. This implies that states with $E = 0$ are doubly degenerate in $k$ space, hence the Fermi surface is not robust in the Wigner-von-Neumann sense.

### 8.3.2 0-flux sector is gapless, but Fermi points are unstable due to nesting

The 0-flux sector has Fermi points when $|J| \leq \frac{1}{2}$. However, they are unstable due to nesting, $\text{spec}(H_k) = -\text{spec}(H_{k+\pi})$ [Chua,Yao,Fiete] (see also discussion in §8.6.2). Upon unit cell doubling, this nesting guarantees spectral symmetry at each $k$ point in the reduced Brillouin zone, yielding a situation similar to the $\pi$-flux case discussed above. Below we show specifically how the Fermi surface degeneracy can be lifted by adding a perturbation in the horizontal bonds with doubled periodicity, shown in Figure 8.2(b). The real space Hamiltonian in the $(A, B, A, B, \cdots)$ basis is

\[
H = H_0 + \Delta H ,
\]

\[
H_0 = \begin{pmatrix}
\sigma_y & iJ \| & -iJ \|

-iJ \| & \sigma_y & \cdots \\
\cdots & \cdots & iJ \|

iJ \| & -iJ \| & \sigma_y
\end{pmatrix}, \quad \Delta H = \epsilon J Q \otimes \begin{pmatrix} 1 \\ \lambda \end{pmatrix}
\]
where $H_0$ is the unperturbed Hamiltonian for $s = 1$ and

$$
Q \equiv \begin{pmatrix}
0 & i & i & i \\
-i & 0 & -i & -i \\
i & 0 & i & i \\
-i & 0 & -i & -i \\
\cdot & \cdot & \cdot & \cdot \\
i & 0 & i & -i \\
-i & -i & i & 0
\end{pmatrix}.
$$  \hspace{1cm} (8.49)

The Fourier transform of $Q$ only couples $k$ with $k + \pi$:

$$
Q_{k,k+q} \equiv \langle k|Q|k+q \rangle = -2i \cos k \cdot \delta_{q,\pi}.
$$  \hspace{1cm} (8.50)

Thus in $k$-space, the full Hamiltonian is diagonalized into $4 \times 4$ blocks,

$$
\tilde{H}_{k;k+\pi} = \begin{pmatrix}
H_k & -2i\epsilon J \cos k \left(\begin{array}{c}
1 \\
\lambda
\end{array}\right)
\\
2i\epsilon J \cos k \left(\begin{array}{c}
1 \\
\lambda
\end{array}\right) & H_{k+\pi}
\end{pmatrix}.
$$  \hspace{1cm} (8.51)

In the unperturbed system, the diagonal blocks $H_k = \sigma_y - 2J \sin k$ are each diagonalized by a $\frac{\pi}{2}$ rotation about the $\hat{x}$ axis, denoted as $D_x$, thus in the unperturbed diagonal basis ($|k,y\uparrow\rangle$, $|k,y\downarrow\rangle$, $|k+\pi,y\uparrow\rangle$, $|k+\pi,y\downarrow\rangle$), the off-diagonal matrix transforms to

$$
\begin{pmatrix}
1 \\
\lambda
\end{pmatrix} \Rightarrow M \equiv D_x \begin{pmatrix}
1 \\
\lambda
\end{pmatrix} D_x^\dagger = \begin{cases}
I, & \lambda = +1 \\
-\sigma_y, & \lambda = -1
\end{cases}.
$$  \hspace{1cm} (8.52)

In the unperturbed system, $|k_F,y\downarrow\rangle$ and $|k_F+\pi,y\uparrow\rangle$ are degenerate at the Fermi level $E_F = 0$ where $k_F$ is one of the Fermi points. To lift the degeneracy, one must have nonzero matrix element between them upon perturbation, which are the off-diagonal elements of $M$. Thus the Fermi points are stable if $\lambda = 1$, but unstable if $\lambda = -1$.

Note that $Q_{r,r+s} = i(-1)^s\delta_{s,\pm 1}$, with $r = 0, 1, \cdots, N - 1$. The Fourier transform is thus

$$
Q_{k,k+q} = \frac{1}{N} \sum_{r,s} e^{i(k+q)s} e^{iqr} Q_{r,r+s} = 2i \cos(k+q) \cdot \sum_r (-e^{iq})^r / N.
$$

The geometric sum vanishes unless $q = \pi$. 

\footnote{Note that $Q_{r,r+s} = i(-1)^s\delta_{s,\pm 1}$, with $r = 0, 1, \cdots, N - 1$. The Fourier transform is thus $Q_{k,k+q} = \frac{1}{N} \sum_{r,s} e^{i(k+q)s} e^{iqr} Q_{r,r+s} = 2i \cos(k+q) \cdot \sum_r (-e^{iq})^r / N$. The geometric sum vanishes unless $q = \pi$.}
**Figure 8.3**: 1D spin ladder with off-diagonal links. The off-diagonal links have strength $K$. $s_1$ and $s_2$ denote the direction of the bonds relative to the arrows. The addition of off-diagonal links breaks nesting and stablizes the Fermi points.

**Stabilizing Fermi points by breaking nesting**

One way to stablize the Fermi points is to add triangular bonds which will break nesting. Consider the modification of Figure 8.3. The $k$-space Hamiltonian is

$$H(k; s_1, s_2) = -2J \sin k \begin{pmatrix} 1 \\ s_1 \end{pmatrix} + (1 + s_2K \cos k)\sigma_y - s_2K \sin k\sigma_x. \quad (8.53)$$

The spectrum is

$$E(k) = \begin{cases} 
-2J \sin k \pm \sqrt{K^2 + 2s_2K \cos k + 1}, & s_1 = +1 \\
\pm \sqrt{4J^2 \sin^2 k + K^2 + 2s_2K \cos k + 1}, & s_1 = -1 
\end{cases}$$

(0 flux per square plaquette)

$$E(k) = \begin{cases} 
\pm \sqrt{4J^2 \sin^2 k + K^2 + 2s_2K \cos k + 1}, & s_1 = -1 
\end{cases}$$

($\pi$ flux per square plaquette)

(8.54)

The $\pi$-flux sector is still inversion symmetric, $\sigma_y H_k \sigma_y = H_{-k}$. It is gapped when $K \neq 1$. When $K = 1$, there is a linear Dirac point at $k = \pi + \cos^{-1}(s_2)$. The 0-flux sector has a gapless phase. Since nesting is broken (e.g., $E(k + \pi) \neq -E(k)$), its Fermi points are stable.
8.3.3 Majorana spectral redundancy and the orientation of Fermi surface

Recall that the vanilla 0-flux spin chain is gapless for $|J| > \frac{1}{2}$. Let $\alpha(\beta)$ denote Majoranas on chain $A(B)$. The full Hamiltonian operator is

$$
\hat{H} = \sum_k (\alpha_{-k}, \beta_{-k}) \begin{pmatrix} H_k \\ \sigma_y - 2J \sin k \end{pmatrix} \begin{pmatrix} \alpha_k \\ \beta_k \end{pmatrix}
$$
(8.55)

$$
= \sum_k (1 - 2J \sin k)c_k^\dagger c_k + (-1 - 2J \sin k)d_k^\dagger d_k
$$
(8.56)

$$
= \sum_k (1 - 2J \sin k)(2c_k^\dagger c_k - 1)
$$
(8.57)

with diagonal basis

$$
c_k = d_{-k}^\dagger = \frac{1}{2}(\alpha_k - i\beta_k).
$$
(8.58)

The spectrum of the matrix $H_k$ is shown in Fig. 8.4(a). Its two bands correspond to $c$ and $d$ fermions in Eq. 8.56, which are not independent but related via Eq. 8.58: an occupied $c_k$ fermion must concur with an empty $d_{-k}$ fermion. The many-body ground state consists of filling both $c$ and $d$ fermions below $E = 0$. Their Fermi surfaces are related. In this picture, unoccupied states do not contribute energy.

Alternatively, one may use Eq. 8.57 without any DOF redundancy, as illustrated in Fig. 8.4(b). The ground state is obtained by filling $c$ fermions with
Figure 8.5: Minimal unit cell bond configuration on the 2D square-octagon lattice. Arrows denote positive link direction. \( s = 1 \) for \( \pi \) flux per square and octagon plaquette, \( s = -1 \) for 0 flux.

\[ E_k = 1 - 2J \sin k < 0 \] (red filled dots), yielding a FS. In this picture, an unoccupied \( c_k \) (i.e. a hole, blue empty dots) also contributes \(-E_k < 0\) to the total energy. There is only one physical band (the red one), its FS has an orientation: states inside are occupied, and states outside are unoccupied.

8.4 2D model: square-octagon lattice

The Kitaev model on a square-octagon lattice has been studied by several authors [78, 4, 37]. Here we only consider the two minimal unit cell cases shown in Figure 8.5 and illustrate some treatments useful for our later generalization to three dimensions. In these cases, the fluxes of the square and octagon plaquettes are always the same.

The \( k \)-space Hamiltonian in the \((A, B, C, D)\) basis is

\[
H(k; J_x, J_y, J_z) = \begin{pmatrix}
  sJ_x \sigma_y & J_y \sigma_y + J_z \alpha(k) \\
  J_y \sigma_y + J_z \alpha^*(k) & -J_x \sigma_y
\end{pmatrix},
\]

\[ \alpha(k) = \begin{pmatrix}
  i e^{i k_y} \\
  -i e^{-i k_z}
\end{pmatrix}, \]

where \( \alpha \) satisfies \( \alpha^*(k) = \alpha^\dagger(k) = -\alpha(-k) \). The ground state is given by \( s = 1 \) where all plaquettes have \( \pi \) flux.
Table 8.1: Symmetry properties of the square-octagon Hamiltonian

<table>
<thead>
<tr>
<th></th>
<th>$k$</th>
<th>$J_x$</th>
<th>$J_y$</th>
<th>$J_z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\tau_x$</td>
<td>$-$</td>
<td>$-s$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$\tau_z$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$\sigma_z$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
<tr>
<td>$\mathcal{K}$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
<td>$-$</td>
</tr>
</tbody>
</table>

### 8.4.1 Fermi surface

A Fermi surface exists if $\det(H_k) = 0$ has solution(s) [78]. Using $\det \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \det \begin{pmatrix} \frac{a}{c} & 0 \\ \chi & \frac{d}{c} \end{pmatrix} = \det(a - bd^{-1}c)\det(d)$ with $X = -d^{-1}c$, we have $H_k = -J_z^2 D(k, s)$ where

$$D(k, s) = J_x^4 + J_y^4 + J_z^4 + 2sJ_x^2J_y^2 - 2J_x^2J_z^2 - 2sJ_x^2J_z^2 \cos(k_x - k_y) - 2sJ_x^2J_z^2 \cos(k_x + k_y).$$

(8.61)

The bounds of $D(k, s)$ are

$$D_{\min}(s) = J_x^4 + J_y^4 + J_z^4 + 2sJ_x^2J_y^2 - 2J_x^2J_z^2 - 2sJ_x^2J_z^2, \quad k = (0, 0) \text{ or } (\pi, \pi),$$

(8.62)

$$D_{\max}(s) = J_x^4 + J_y^4 + J_z^4 + 2sJ_x^2J_y^2 + 2J_x^2J_z^2 + 2J_y^2J_z^2, \quad k = (0, \pi) \text{ or } (\pi, 0).$$

(8.63)

Existence of Fermi surface requires $D_{\min} \leq 0 \leq D_{\max}$.

In the $\pi$-flux sector, $s = 1$ and $D_{\min} = (J_x^2 + J_y^2 - J_z^2)^2 \geq 0$. The gap vanishes only when $J_x^2 + J_y^2 = J_z^2$ and the Fermi points are $k = (0, 0)$ and $\pi, \pi$.

In the $0$-flux sector, $s = -1$. $D_{\max} = (J_x^2 - J_y^2 - J_z^2)^2 + 4J_x^2J_z^2 \geq 0$.

Existence of Fermi surface then requires $J_x^4 + J_y^4 + J_z^4 \leq 2(J_x^2J_y^2 + J_y^2J_z^2 + J_z^2J_x^2)$. This condition is symmetric in the $x, y, z$ indices. Rearranging, one gets

$$(J_x^2 + J_y^2 - J_z^2)^2 \leq 4J_x^2J_z^2 \quad \Rightarrow \quad |J_x| - |J_y| \leq |J_z| \leq |J_x| + |J_y|$$

(8.64)

and cyclic in $x, y, z$. In other words, $|J_x|, |J_y|$ and $|J_z|$ form a triangle.

### 8.4.2 Spectral symmetry and stability of Fermi surface

The symmetry properties of the Hamiltonian are shown in Tab. 8.1, where $\mathcal{K}$ denotes complex conjugation. For example, the first line means $\tau_x H(k, J_x, J_y, J_z) \tau_x = $
\[ H(-k, -sJ_x, J_y, J_z). \]

In the \( \pi \)-flux sector, \( s = 1 \). The sign of each \( J_{x,y,z} \) can be individually flipped without changing the energy spectrum: \( J_x \) by \( \tau_x \sigma_z K \), \( J_y \) by \( \tau_y K \), and \( J_z \) by \( \tau_y K \) (\( \tau_y \) equivalent to \( \tau_x \tau_z \)). In particular, we have the following inversion symmetry,
\[ \tau_y \otimes \sigma_z : H(k) \longrightarrow H(-k). \quad (8.65) \]

Then similar to the the 1D case, the \( \pi \)-flux sector should in general be gapped, and even when there is a Fermi surface, it is unstable in the Wigner-von-Neumann sense. This agrees with the gapless condition in §8.4.1.

In the 0-flux sector, \( s = -1 \). There is no combination in Eq. 8.1 which will map \( H(k) \) to \( +H(-k) \), hence no \( k \)-local spectral symmetry. This implies that in a gapless phase, the Fermi level at any \( k \) is non-degenerate. However, the Fermi surface will still be \textit{unstable due to nesting} by \( \pi \equiv (\pi, \pi) \). This is because shifting by \( \pi \) is equivalent to flipping \( J_z \), which, when \( s = -1 \), can be mapped by \( \tau_z \) to \( H(-k) = -H^*(k) \). Thus similar to the 1D 0-flux case, upon enlarging the unit cell along the \( \pi \) direction, the energy spectrum has \( k \)-local spectral symmetry, and Fermi level at any \( k \) is doubly degenerate, which can be lifted by a perturbation coupling the two degenerate levels.

### 8.5 \( N = 2k \): Clifford algebra and Gamma matrices

Kitaev models on lattices with coordination number 3 can be formulated using Pauli matrices, which represent rank-2 Clifford algebra (CA). Higher coordination numbers require the use of higher rank CAs and their accompanying generalized Gamma matrices. In this section we summerize a systematic way to construct Gamma matrices associated with rank-2\( k \) CAs with integer \( k \).

An \( N = 2k \)-dimensional Clifford algebra (CA) has \( 2k \) anticommuting generators \( \{ \gamma^a \} \),
\[ \{ \gamma^a, \gamma^b \} = 2\delta^{ab}, \quad a, b = 1, 2, \cdots, 2k. \quad (8.66) \]

Their product gives the pseudoscalar
\[ \gamma^{2k+1} \equiv (-i)^k \gamma^1 \gamma^2 \cdots \gamma^{2k}, \quad (\gamma^{2k+1})^2 = 1. \quad (8.67) \]
A systematic way exists to generate their matrix representation,

\[ \gamma^{1} = \sigma^{x} \otimes I \otimes I \otimes \cdots \otimes I, \quad \gamma^{3} = \sigma^{z} \otimes \sigma^{x} \otimes I \otimes \cdots \otimes I, \quad \cdots \]  
(8.68)

\[ \gamma^{2} = \sigma^{y} \otimes I \otimes I \otimes \cdots \otimes I, \quad \gamma^{4} = \sigma^{z} \otimes \sigma^{y} \otimes I \otimes \cdots \otimes I, \quad \cdots \]  
(8.69)

\[ \gamma^{2k+1} = \sigma^{z} \otimes \sigma^{z} \otimes \sigma^{z} \otimes \cdots \otimes \sigma^{z}. \]  
(8.70)

The \(2k+1\) \(\gamma\)-generators can also be represented in terms of \(2k+2\) Majoranas,

\[ \gamma^{a} = i\theta^{0}\theta^{a}, \quad a = 1, 2, \cdots, 2k+1 \]  
(8.71)

where \(\{\theta\}\) are anticommuting Majoranas,

\[ \{\theta^{\alpha}, \theta^{\beta}\} = 2\delta^{\alpha\beta}, \quad \alpha, \beta = 0, 1, \cdots, 2k+1. \]  
(8.72)

Eq. 8.67 then imposes a physical restriction,

\[ \theta^{0}\theta^{1}\cdots\theta^{2k+1} \equiv i^{k-1}. \]  
(8.73)

Example: the Pauli matrices constitute a basis of the 2-dimensional CA (i.e. \(k = 1\)),

\[ \{\sigma^{i}, \sigma^{j}\} = 2\delta_{ij} \quad \text{for} \quad i, j = x, y, \]  
(8.74)

\[ \sigma^{z} = (-i)^{k=1}\sigma^{x}\sigma^{y}, \]  
(8.75)

\[ \sigma^{x,y,z} = i\theta^{0}\theta^{x,y,z}, \quad \theta^{0}\theta^{x}\theta^{y}\theta^{z} \equiv 1. \]  
(8.76)

### 8.6 3D model with one Majorana per site: the octahedron-cubic lattice (OCL)

The OCL can be obtained by replacing each vertex in a cubic lattice with an octahedron (Fig. 8.6), and has a coordination number of 5, thus a Kitaev model is to be written in 6-dimensional Clifford algebra \(\Gamma^{a}, \ a = 1, \cdots, 5\), and \(\Gamma^{ab} = i[\Gamma^{a}, \Gamma^{b}]/2\). It can be represented by introducing 6 Majorana operators \(\theta^{0}, \cdots, \theta^{5}\) and writing

\[ \Gamma^{a} = i\theta^{0}\theta^{a} \quad (a = 1, 2, 3, 4, 5), \quad \mathbb{J} \equiv i\theta^{0}\theta^{1}\theta^{2}\theta^{3}\theta^{4}\theta^{5} = -1 \]  
(8.77)
where the second constraint follows from $\Gamma^1\Gamma^2\Gamma^3\Gamma^4\Gamma^5 = -1^6$ and can be enforced by applying the projector $P = (1 - J)/2$ to states written in the Majorana representation.

### 8.6.1 Octahedron flux configuration

An octahedron has eight triangular plaquettes, as shown in Figure 8.6. The positive direction of the flux through a plaquette is defined as going out of the octahedron. It is easy to verify that the product of all eight plaquettes must be positive, thus the number of plaquettes with negative flux is even. Flux configurations related by the octahedron symmetry are equivalent. There are ten inequivalent flux configurations as shown in Figure 8.7: class $A$ with no negative plaquette, class $B_1$ with two negative plaquettes, and class $C_1$ with four negative plaquettes.

By inverting all plaquettes (equivalent to time-reversal operation which inverts all bonds), one obtains configurations with more than four negative plaquettes; numerically it is the same as reverting the sign of all hopping parameters.

Each flux configuration can be realized by various gauges (directions on links), their respective real space Hamiltonian being related through gauge transformations. Note that equivalent gauges do not necessarily render the same $k$-space Hamiltonian, i.e., the spectrum $E(k)$ may be different—they are equivalent in the sense that the full set of eigenvalues is the same. Figure 8.7 shows a specific gauge configuration for class $A$; the gauge fields of other classes are obtained by flipping the red links relative to their values in the $A$ class.

---

$^6$This can be checked using the matrix representation $\Gamma^{(1,2,3,4,5)} = (\sigma_x, \sigma_y, \tau_y\sigma_z, \tau_y\sigma_z, \tau_z\sigma_z)$. 

**Figure 8.6:** Octahedron and its planar representation
Figure 8.7: Octahedron flux classification. White triangles have positive flux (out of the page), purple ones have negative flux. Multiplicity of each class by the octahedron symmetry group is shown in the label, and sum to 99, agreeing with \( \binom{8}{0} + \binom{8}{2} + \binom{8}{4} = 99 \). One possible link configuration for \( A \) is shown. Other flux configurations can be implemented by flipping the red links relative to their values in the \( A \) class. Note that since there are 8 plaquettes and 1 constraint (that their product be +1), one only need the ability to tune \( 8 - 1 = 7 \) links in order to realize all flux configurations. Here we have frozen the rest \( 12 - 7 = 5 \) links in all classes, \( u_{a\bar{z}} = u_{xx} = 1 \) where \( a \in \{x, \bar{x}, y, \bar{y}\} \). Time-reversal partners are obtained by flipping the sign of all triangular plaquettes and are not shown here.

8.6.2 Minimal unit cell (MUC)

The magnetic unit cell is identical to the structural unit cell if all octahedrons have the same gauge configuration, and cubic bonds are set as positive in the \( \hat{x}, \hat{y} \) and \( \hat{z} \) direction. The three cubic (i.e. octagon) fluxes are entirely determined and identical to the three equator fluxes of the octahedron. We use the gauge field setup in Figure 8.8 to implement a given flux configuration.

Ground state energy of different flux classes

The ground state energy of each flux class can be obtained by Equation 8.33 with all \( N_n = 0 \), i.e., \( E_{GS} = -2 \sum_{n=1}^{M} \varepsilon_n \). We adopt the parameterization \( (J_o, J_c) = (\cos \eta, \sin \eta) \), where \( J_o \) and \( J_c \) are strength of octahedron and cubic links, respectively. Changing the sign of both \( J_o \) and \( J_c \) is equivalent to inverting all bonds and does not change the spectrum of a Majorana Hamiltonian. Changing the sign of \( J_c \) alone does not change any flux in the MUC scheme. Thus one can restrict \( \eta \in [0, \frac{\pi}{2}] \). The ground state energy of different fluxes as a function of
**Figure 8.8**: Flux to link mapping for a single octagon. Triangular faces are numbered and $f_i$ denotes the flux through face $i$. Five links are frozen, $u_{az} = u_{xz} = 1$ with $a \in \{y, \bar{y}, z, \bar{z}\}$. The value of the rest of the links can be determined by $\{f_i\}$. The link $u_{xz}$ can be written in two ways as shown, expressing the fact that the product of all $f_i$ is +1.

$\eta$ is shown in Figure 8.9. It is interesting to note that in all range of $\eta$, the extremal energy is given by inversion-symmetric fluxes, $A$ and $C_3$ for minimal and $C_5$ for maximal. Here inversion-symmetric means the flux configuration (but not necessarily the gauge fields) is invariant upon lattice inversion $(x, y, z) \leftrightarrow (\bar{x}, \bar{y}, \bar{z})$.

**Phase diagram of different fluxes**

Figure 8.10 plots for each class the energy gap (red lines) and the number of levels crossed by $E = 0$. A gapless phase with no level crossing is analogous to a semimetal.

**Octahedron limit $J_o \gg J_c$ ($\eta = 0$) and bulk zero modes**

In the octahedron limit, the Hamiltonian reduces to that of a single octahedron,

$$H_o = \begin{pmatrix}
0 & v_{xy} & v_{xz} \\
v_{xy}^\dagger & 0 & v_{yz} \\
v_{xz}^\dagger & v_{yz}^\dagger & 0
\end{pmatrix}$$

(8.78)

which is written in the basis $(x, \bar{x}, y, \bar{y}, z, \bar{z})$. The purely imaginary offdiagonal blocks are determined by the gauge fields, shown in Table 8.2.
**Figure 8.9:** MUC ground state energy of different flux classes relative to class $A$. $(J_{\text{oct}}, J_{\text{cub}}) = (\cos \eta, \sin \eta)$. Depending on $\eta$, either $A$ or $C_3$ has lower energy. Inset shows the unsubtracted energy. One can see that the difference in the cubic limit ($\eta \to \pi/2$) is extremely small. It is interesting to note that the extremal cases ($A, C_3$ and $C_5$) all have inversion-symmetric flux configuration. $20 \times 20 \times 20$ octahedrons are used.

**Table 8.2:** Offdiagonal blocks of the OCL Hamiltonian in the octahedron limit.

<table>
<thead>
<tr>
<th></th>
<th>$A$</th>
<th>$B_1$</th>
<th>$B_2$</th>
<th>$B_3$</th>
<th>$C_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_{xy}$</td>
<td>$i - i\sigma_x$</td>
<td>$i - i\sigma_x$</td>
<td>$i - i\sigma_x$</td>
<td>$i - i\sigma_x$</td>
<td>$i - i\sigma_x$</td>
</tr>
<tr>
<td>$v_{yz}$</td>
<td>$-i\sigma_z - \sigma_y$</td>
<td>$i - \sigma_y$</td>
<td>$i - \sigma_y$</td>
<td>$i - \sigma_y$</td>
<td>$i + i\sigma_x$</td>
</tr>
<tr>
<td>$v_{xz}$</td>
<td>$i + i\sigma_x$</td>
<td>$i + i\sigma_x$</td>
<td>$i - \sigma_y$</td>
<td>$i - i\sigma_x$</td>
<td>$i + i\sigma_x$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>$C_2$</th>
<th>$C_3$</th>
<th>$C_4$</th>
<th>$C_5$</th>
<th>$C_6$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$v_{xy}$</td>
<td>$i - \sigma_y$</td>
<td>$i + \sigma_y$</td>
<td>$-i\sigma_z - i\sigma_x$</td>
<td>$-i - i\sigma_x$</td>
<td>$i + i\sigma_x$</td>
</tr>
<tr>
<td>$v_{yz}$</td>
<td>$i + i\sigma_x$</td>
<td>$i - \sigma_y$</td>
<td>$-i\sigma_z - \sigma_y$</td>
<td>$i + i\sigma_x$</td>
<td>$-i\sigma_z - \sigma_y$</td>
</tr>
<tr>
<td>$v_{xz}$</td>
<td>$i + i\sigma_x$</td>
<td>$i - \sigma_y$</td>
<td>$i - i\sigma_y$</td>
<td>$i + i\sigma_x$</td>
<td>$i + i\sigma_x$</td>
</tr>
</tbody>
</table>
Figure 8.10: Gap and level crossing in OCL with MUC flux. Gap is plotted in red and the number of level crossing in blue. In a gapless phase, nonzero level crossing indicates the existence of a Fermi surface whereas zero level crossing is analogous to a semimetal phase. Note that the “minigaps” of $C_3$ actually belong to a semimetal phase, with nodes occurring along $k_y = k_z = -k_x$ using the gauge of Figure 8.7, which is numerically uncovered. $20 \times 20 \times 20$ octahedrons are used.
Note that several of the classes have two of the three blocks identical. Assume the three blocks are \(a, b, b\) (order doesn’t matter), then by adding or subtracting rows/columns of blocks, which leaves the determinant invariant, one obtains

\[
M = \begin{pmatrix}
0 & a & b \\
\dagger a & 0 & b \\
\dagger b & \dagger b & 0
\end{pmatrix} \implies \begin{pmatrix}
0 & a & b \\
\dagger a & -a & 0 \\
\dagger b & \dagger b & 0
\end{pmatrix} \implies \begin{pmatrix}
\dagger a & -a-a^\dagger & 0 \\
\dagger b & 0 & 0
\end{pmatrix}.
\] (8.79)

A sufficient condition for \(\det M = 0\) is that the second and third block-columns are linearly dependent, viz.,

\[
\det \begin{pmatrix}
a & b \\
-a-a^\dagger & 0
\end{pmatrix} = \det b \cdot \det(a + a^\dagger) = 0
\] (8.80)

One can check that \(a + a^\dagger = 0\) for \(B_2, C_1, C_5\), and \(\det b = 0\) for \(B_3, C_2, C_6\), thus they have zero modes. The other four classes have no zero modes, although \(C_3\) does have two identical blocks\(^7\). See red lines at the \(\eta = 0\) point in Figure 8.10.

**Cubic limit \(J_o \ll J_c (\eta = \frac{\pi}{2})\) and constant surface zero modes**

In the cubic limit, the Hamiltonian reduces to decoupled hopping in the three directions,

\[
H_c(k_x, k_y, k_z) = J_c \begin{pmatrix}
h(k_x) \\
h(k_y) \\
h(k_z)
\end{pmatrix},
\] (8.81)

\[
h(k_i) = \begin{pmatrix}
\dagger e^{-i \eta} \\
e^{i \eta}
\end{pmatrix} \equiv B(k_i) \cdot \sigma.
\] (8.82)

The bulk energies are \(\varepsilon = \pm J_c\) each with three-fold degeneracy. As \(k_i\) goes from 0 to \(2\pi\), \(B(k_i)\) winds around the origin in the \(xy\) plane. Thus surface energy in all three directions is zero.

**Semimetal phase in \(C_3\)**

While \(B_2, B_3, C_2\) and \(C_3\) all show signs of a semimetal phase, \(C_3\) is qualitatively different. All but \(C_3\) have their prospective semimetal phase derived from

\(^7\)I wonder if it can be proved that \(\det a \cdot \det b = 0\) is both sufficient and necessary.
their octahedron limit zero modes, see §8.6.2. As \( \eta \) shifts away from zero, the two zero bands split as well as develop asymmetric band fluctuations. When the fluctuations exceed the band splitting, the number of level crossing \( E = 0 \) becomes nonzero (for symmetric band fluctuations it remains zero) and the system enters a metallic phase\(^8\). For \( C_3 \), the semimetal phase intervenes between the two gapped phases of the octahedron and cubic limits.

Using the gauge fields of Figure 8.8, one obtains \( \det H(k) \) on three special \( k \)-space planes\(^9\)

\[
\det H(q, k, k; j) = \det H(k, q, -k; j) = \det H(k, -k, q; j) = -|f(k, q; j)|^2 \quad (8.83)
\]

where \( j \equiv J_c/J_o = \tan \eta \) is positive (recall \( \eta \in [0, \frac{\pi}{2}] \)), and

\[
f(k, q; j) = j^3 - 2i \sin k \cdot j^2 - 2j - 4e^{iq}(j \cos k + 1) \ . \quad (8.84)
\]

The overall sign of \( \det H \) reflects the spectral symmetry at each \( k \) (three levels below zero). The gap closes when \( f = 0 \), viz.,

\[
e^{iq} = j^2 - 2i \sin k \cdot j - 2j^2 - 2j - 4e^{iq}(j \cos k + 1) \ . \quad (8.85)
\]

Imposing unimodularity on the RHS yields a quadratic equation of \( \cos k \) with solutions \( \cos k = c_j(j) \),

\[
c_1(j) = -\frac{j^4 + 2j^2 + 8}{2j^3 + 8j} \quad , \quad c_2(j) = \frac{j^2 - 2}{2j} \ . \quad (8.86)
\]

Lower bounding \( c_1(j) \geq -1 \) implies \( (j - 2)(j^3 + 2j - 4) \leq 0 \) which restricts \( j \), for the \( c_1 \) branch, to

\[
j_0 \leq j \leq j_\pi \ , \quad j_0 \simeq 1.180 \ (\eta \simeq 0.276\pi) \ , \quad j_\pi = 2 \ (\eta \simeq 0.352\pi) \ . \quad (8.87)
\]

The gap closure associated with \( c_1(j) \) thus occurs at \( k = (q_1, k_1, k_1), (k_1, q_1, -k_1) \) and \( (k_1, -k_1, q_1) \), which are three-fold symmetric about the cubic diagonal \( -k_x = k_y = k_z \), with

\[
k_1(j) = \pm \cos^{-1}[c_1(j)] \quad , \quad q_1(j) = \pi + \text{arg}(j^2 - 2i \sin k_1 \cdot j - 2) \ . \quad (8.88)
\]

\(^8\)According to this picture though, the system should be in an insulating phase before the fluctuation exceeds band splitting. This needs further numerical study.

\(^9\)Form of \( \det H \) obtained from GNU Maxima.
Figure 8.11: $k$-space location of $E = 0$ nodes. In (a), solid colored lines show the $k$-space trajectory of nodes arising from the $c_1$ branch (see Equation 8.86), dashed lines show their projection onto the 2D Brillouin zones. The red, blue, and light blue branches lie on $k_y = k_z$, $k_x = -k_z$, and $k_x = -k_y$ planes, respectively, as can be seen from their 2D projections. (b) shows the $j$ dependence of $k_1$, $q_1$ (of the $c_1$ branch) and $k_2$ (of the $c_2$ branch). It also serves as a topological phase diagram of the effectively 2D system $H(k_x, k_y)$ at fixed $k_z$ and $j$, see text.

These are plotted in Figure 8.11b. Each $j$ gives rise to two opposite $k_1$, resulting in six gap collapsing points in general, but the opposing two are equivalent due to the Majorana nature. For $j = j_0$, they form three degenerate pairs at $(0, \pi, \pi)$, $(\pi, 0, \pi)$ and $(\pi, \pi, 0)$. For $j = j_\pi$, they coincide at $(\pi, \pi, \pi)$. The trajectory these nodes are shown in Figure 8.11a.

For the $c_2(j)$ branch, $|c_2(j)| \leq 1$ yields

$$j^- \leq j \leq j^+ \ , \quad j^\pm = \sqrt{3} \pm 1 \ (0.201\pi \lesssim \eta \lesssim 0.388\pi). \quad (8.89)$$

Substituting Equation 8.86 in 8.85 gives $q = -k$. The associated gap closure thus occurs along the line $-k_x = k_y = k_z \equiv k_2$ where

$$k_2(j) = \pm \cos^{-1} [c_2(j)]. \quad (8.90)$$

Each $j$ yields an opposing pair of $k_2$, which are equivalent due to the Majorana nature. For $j = j^-$, the gap closes at $(\pi, \pi, \pi)$. For $j = j^+$, the gap closes at $(0, 0, 0)$.

The topological property of the system can be considered in the following way. At fixed $j$ and $k_z$ (similar analysis applies for fixed $k_x$ or $k_y$), the Hamiltonian
is effectively two-dimensional and can be characterized by a Chern number. Figure 8.11b is then a phase diagram in the $k_z - j$ space, where the lines of $k_1, q_1$ and $k_2$ are the phase boundaries on which the gap of $H(k_x, k_y)$ vanishes. There are two $E = 0$ nodes at $k_z = k_1$, and one at $k_z = q_1$ and $k_2$ each, thus the Chern number changes by 1 across $q_1$ and $k_2$, and by 2 across $k_1^{10}$. There are thus five phases, $N$ with trivial band structure, $C^+_1;_2$ with a Chern number $C = 1$, $C^-$ with $C = -1$, and $S$ which is gapped in the bulk but has isolated midgap surfaces states detached from the bulk. The Wannier flows in the $x$ direction for all five phases are shown in Figure 8.12. Note the qualitative difference of $N$ and $S$: in $N$, all three Wannier levels fluctuate around $\gamma = 0$ whereas in $S$, one of them fluctuates around $\pi$, in agreement with the occurrence of the isolated midgap surface states.

**Metallic phases**

Figs. 8.13 and 8.14 show typical Fermi surface (FS) geometry of the 8 metallic phases. The Majorana nature guarantees that the $3^{rd}$ and $4^{th}$ bands are related via $\varepsilon_3(k) = -\varepsilon_4(-k)$, thus their respective FSs form inversion pairs. For $C_1$ and $C_5$, the two band FSs are cuboids, allowing a (pairing) nesting momentum $q = (\pi, \pi, \pi)$ to connect them. $B_3$ has a (normal) nesting of $q = (\pi, \pi, \pi)$ for both branches. All classes also exhibit certain reflection symmetry, some within each band (“self-symmetric”), some between the two bands. These are noted in their captions.

Now consider the FS stability. In the language of Ref. [15], the blue and yellow FSs in Fig. 8.13 are ($k$-inverted) duplicates of each other due to the Majorana nature and one should think of just one independent branch, say the yellow one. Ref. [15] considered two types of FS nesting that would lead to FS instability: (i) normal nesting, where an independent branch itself exhibits nesting, e.g. $\varepsilon_3(k) = \pm \varepsilon_3(k + q)$, and (ii) pairing nesting, where there is a nesting between the two branches, e.g. $\varepsilon_3(k) = \pm \varepsilon_4(k - q)$ which implies $\varepsilon_3(k) = \mp \varepsilon_3(q - k)$. As noted earlier, $C_1$ and $C_5$ exhibit a pairing nesting, and $B_3$ a normal nesting.

\[^{10}\text{Actually, assuming each occurrence of band touching exchanges a Chern number } C \text{ of magnitude one, then one can only conclude that crossing } k_1 \text{ changes } C \text{ by 0 or 2.}\]
Figure 8.12: $x$-direction Wannier flow of MUC $C_3$ class
(a) $B_1$, $\eta = 0.28\pi$, symmetric about $k_y = k_z$

(b) $B_2$, $\eta = 0.26\pi$, symmetric about $k_x = -k_y$

(c) $B_3$, $\eta = 0.18\pi$, symmetric about $k_x = k_z = k_y$, normal nesting $q = (\pi, \pi, \pi)$

(d) $C_1$, $\eta = 0.26\pi$, self-symmetric about $k_y$, pairing nesting $q = (\pi, \pi, \pi)$

**Figure 8.13**: Fermi surfaces of OCL metallic phases. See description of Fig. 8.14
Figure 8.14: Fermi surfaces of OCL metallic phases. The 3rd and 4th bands are related by $\varepsilon_3(k) = -\varepsilon_4(-k)$ due to the Majorana nature. Fermi surface of the 3rd band, $\varepsilon_3(k) = 0$, are shown in yellow, and that of the 4th band in blue. Energy gradient (lower to higher) points from the wireframed side to the unframed side. Note that the two band FSs in $C_1$ (Fig. 8.13d) and $C_5$ (Fig. 8.14c) are related by nesting. All FSs exhibit certain reflection symmetry, either within the same band (“self-symmetric”) or between the two bands, as noted in the captions. See text for discussion on stability. $k$ space mesh is $50 \times 50 \times 50$. 

(a) $C_2$, $\eta = 0.28\pi$, self-symmetric about $k_x = k_y$

(b) $C_4$, $\eta = 0.34\pi$, self-symmetric about $k_x = 0$

(c) $C_5$, $\eta = 0.28\pi$, self-symmetric about $k_x = k_y, k_y = k_z, k_x = k_z$, pairing nesting

(d) $C_6$, $\eta = 0.28\pi$, symmetric about $k_x = -k_y$
They are therefore unstable against certain density waves along the $(1,1,1)$ direction. The cuboid shape of the FS of $C_1$ and $C_5$ however may have derived from the isotropy in the gauge field strengths and could be broken by using anisotropic strengths, thereby stabilizing the FS. The rest flux classes show neither type of nesting, and are therefore stable in the sense of Ref. [15].

Note that the FS of the 2D square-octagon model (§8.4) is unstable due to pairing nesting. The FS of the OCL lattice, when viewed along one of the three $k$ axes (i.e. projected onto any 2D BZ), would also exhibit pairing nesting. One may thus say that the 3rd dimension stabilizes the FS.

As we shall see in §8.6.3, Lieb’s Theorem demands a reflection-symmetric flux configuration in the ground state. This is possibly related to the FS reflection symmetry seen in Fig. 8.13. For example, assume the yellow and blue FSs are reflection copies of each other. Upon unit cell doubling in all three directions, the BZ is triple-halved, which brings the blue FS to overlap with yellow one. They are not degenerate as of yet (because of the lack of nesting). Now apply flux-reflection-symmetrization. This possibly morphes both FSs—originally reflection copies of each other—into the same shape, i.e. a doubly degenerate FS (typical of semimetallic band structure). This is consistent with Fig. 8.17 where all muc metallic phases becomes semimetallic. Such a degeneracy, if indeed exists, is unstable in the Wigner-von-Neumann sense. However, since this should not be deemed a perturbation (it involving a global change in flux configuration), it does not affect the conclusion that the FSs of the un-symmetrized cases are stable.

### 8.6.3 Octuple unit cell (OUC)

For a reflection symmetric lattice, Lieb’s Theorem demands the ground state to have (i) a reflection-symmetric flux configuration with respect to all reflection planes, and (ii) a $\pi$ flux through all plaquettes bisected by any reflection plane. The OCL lattice is reflection symmetric about planes that bisect the square (i.e. octagon) plaquettes without cutting through any octahedron, thus its ground state must have a flux (i.e. gauge potential) unit cell consisting of $2 \times 2 \times 2$ structural unit cells, in which the $6 \times 4 = 24$ square plaquettes all have $\pi$ flux and the $2^3 = 8$
octahedrons belong to the same flux class but with their relative orientations fixed by reflection. Note that the square flux can only be $0\pi$ in the MUC scheme.

The cubic $\pi$ flux can be implemented in a magnetic (i.e. gauge field) unit cell consisting of $2 \times 2 \times 2$ octahedra, with cubic-link gauge fields shown in Fig. 8.15. The relative displacement of an octahedron from the base unit (which is labeled by 0) can be used to label the octahedron, and determines its gauge fields by applying corresponding reflections to that of the base unit. Thus the octahedron labeled by $xy$, for example, is located $(1, 1, 0)$ away from the base unit, and its gauge fields are obtained by applying both $x$ and $y$ reflections on the base unit.

The PBC Hamiltonian is

$$H_0 = \begin{pmatrix}
0 & h_0 & v_x & v_y & v_{xy} & v_z & z & zx & zy & zxy \\
0 & h_x & v_x & v_y & v_{xy} & v_z & z & zx & zy & zxy \\
0 & v_y & h_y & -v_x & v_{xy} & v_z & z & zx & zy & zxy \\
v_y & v_{xy} & -v_x & h_{xy} & v_z & z & zx & zy & zxy & zxy \\
h_z & v_x & -v_y & v_{xy} & -v_y & -v_x & h_{xy} & v_z & z & zx & zy & zxy \\
-h_z & -v_x & v_y & -v_x & h_{xy} & v_z & z & zx & zy & zxy & zxy & zxy \\
-v_z & v_x & -v_y & v_{xy} & -v_y & -v_x & h_{xy} & v_z & z & zx & zy & zxy & zxy \\
v_{z} & v_{z} & v_{z} & v_{z} & v_{z} & v_{z} & v_{z} & v_{z} & v_{z} & v_{z} & v_{z} & v_{z} & v_{z} \\
\end{pmatrix}$$

(8.91)

where blocks with a minus sign correspond to red links in Fig. 8.15. Here, $h_r$ is the local hopping on the octahedron $r$ with $r \in \{0, x, y, xy, z, zx, zy, zxy\}$, and is proportional to the octahedron hopping strength $J_o$, whereas

$$v_a = iJ_c \left[ |a\rangle\langle \bar{a}| - e^{-ik_a} |\bar{a}\rangle\langle a| \right], \quad a = x, y, z$$

(8.92)
Figure 8.16: OUC ground state energy of different flux classes relative to class $A$. $(J_{\text{oct}}, J_{\text{cub}}) = (\cos \eta, \sin \eta)$. Now $A$ always has the lowest energy. Inset shows the unsubtracted energy. Comparing the inset here with Fig. 8.9, one can see that imposing Lieb’s theorem only modifies energy by very small amount, but it is sufficient to stabilize flux $A$. It is interesting to note that the extremal cases ($A, C_5$) all have inversion-symmetric flux configuration. $20 \times 20 \times 20$ octahedrons are used ($10^3$ OUCs).

are hoppings on the cubic links in the $a$ direction, with $|a\rangle$ and $|\bar{a}\rangle$ denoting the octahedron basis.

Fig. 8.17 plots bulk phase diagrams of the ten octahedron fluxes a la Fig. 8.10. There is no truely metallic phase where FS is non-degenerate (which would have a nonzero blue line). MUC metallic and semimetallic phases now develop minigaps, but without anything interesting (in terms of Wannier winding) happening in them (checked, not shown). Note that semimetallic FSs are always unstable in the Wigner-von-Neumann sense, and should degrade to line/point nodes a la Weyl semimetals.

### 8.7 3D model with two Majoranas per site: the pyrochlore lattice

The pyrochlore lattice has a coordination number of 6, thus a Kitaev-like model needs at least 7 majoranas, where 6 of them are to be fused as $Z_2$ fields
Figure 8.17: Gap and level crossing in OCL with OUC bond configuration. Gap is plotted in red, and the number of level crossings in blue. In a gapless phase, nonzero level crossing indicates the existence of a Fermi surface whereas zero level crossing is analogous to a semimetal phase. $20 \times 20 \times 20$ octahedrons are used ($10^3$ OUCs). These should be compared with Fig. 8.10.
and the one left as the hopping $\theta^0$. For a Gamma matrix model, the number of majoranas has to be even, hence one needs at least 8 majoranas corresponding to $k = 3$ (cf. §8.5). The fermion sector thus consists of two freely hopping majoranas with hybridization,

$$H = i \sum_{\langle ij \rangle, a=1,2} J_{ij}^a u_{ij} \eta_i^a \eta_j^a + iK \sum_i \left[ \eta_i^1 \eta_i^2 - \eta_i^2 \eta_i^1 \right],$$  \hspace{1cm} (8.93)

where $i, j$ label pyrochlore lattice sites, $a = 1, 2$ denote species of the hopping majoranas $\eta^a$, and $u_{ij} = \pm 1$ the $Z_2$ gauge field. $J$ and $K$ are the hopping and hybridization strengths.

Geometrically, the pyrochlore lattice consists of two sets of opposite-pointing tetrahedra. Its structural unit cell is an up-pointing (say) tetrahedron. The set of edges connecting these unit cells form the down-pointing tetrahedra. Alternatively one can think of it as composed of four Kagome planes arranged in a tetrahedral fashion, cf. Fig. 8.18. The basis vectors $\mathbf{a}_{1,2,3}$ are along $da$, $db$ and $dc$, respectively.

Since a supertetrahedron can be embedded in a cube with its edges being the cubic face diagonals, one can choose the basis vectors as

$$\mathbf{a}_1 = (0,1,1) \quad , \quad \mathbf{a}_2 = (1,0,1) \quad , \quad \mathbf{a}_3 = (1,1,0),$$  \hspace{1cm} (8.94)
with corresponding reciprocal lattice basis vectors

\[
\begin{align*}
\frac{b_1}{\pi} &= (-1, 1, 1), \\
\frac{b_2}{\pi} &= (1, -1, 1), \\
\frac{b_3}{\pi} &= (1, 1, -1),
\end{align*}
\tag{8.95}
\]

one can check that \( \mathbf{a}_i \cdot \mathbf{b}_j = 2\pi \delta_{ij} \). In this coordinate system, Fig. 8.18 is viewed along the \([1, 1, 1]\) direction.

The first BZ can be constructed in the following way. In the rectilinear \( k \)-space, the origin \( \mathbf{G} = (0, 0, 0) \) has eight nearest neighbor points, \((\pm 1, \pm 1, \pm 1)\pi\) with uncorrelated signs. In terms of \( \mathbf{b}_i \), these are \( \pm \mathbf{b}_i \) and \( \pm \sum_i \mathbf{b}_i \). The first BZ consists of vectors \( \mathbf{k} \) inside the cube \([-\pi, \pi] \times [-\pi, \pi] \times [-\pi, \pi]\), which are closer to the origin than to the eight bounding points.\(^{11}\)

The pyrochlore lattice is highly frustrated because of the abundance of triangular plaquettes, and there is no Lieb’s theorem for the ground state flux configuration due to the lack of reflection planes. Thus we shall limit ourselves to a minimal flux unit cell, i.e. identical to a structural unit cell. The link unit cell (i.e. magnetic unit cell) could be bigger.

### 8.7.1 Tetrahedron flux configuration

The product of the four fluxes on each tetrahedron is +1, hence the number of negative fluxes must be even. There are thus 6 inequivalent tetrahedron flux configurations, shown in Fig. 8.19. All other configurations are related by time-reversal (i.e. flipping all links on both tetrahedron) or tetrahedral symmetry (i.e. attaching the two tetrahedra at a different vertex).

### 8.7.2 Hexagonal flux configuration

When the flux (not necessarily the link) configurations of all elementary up-pointing tetrahedra are identical, the product of the four hexagonal fluxes on an up-pointing super-tetrahedron is identical to the product of triangular fluxes on an up-pointing elementary tetrahedron, i.e. there is an even number of outgoing

\(^{11}\)The next nearest neighbors to the origin are \( \pm (\mathbf{b}_2 + \mathbf{b}_3) = \pm (2\pi, 0, 0) \), \( \pm (\mathbf{b}_3 + \mathbf{b}_1) = \pm (0, 2\pi, 0) \), and \( \pm (\mathbf{b}_1 + \mathbf{b}_2) = \pm (0, 0, 2\pi) \). All points inside the bounding cube are closer to the origin than to these second nearest neighbors.
Figure 8.19: Flux and gauge field configuration of a pair of corner-sharing tetra-
hedra. The vertices $a, b, b', c, c'$ all lie in the same plane. $d$ is above plane and $d'$ is
beneath the plane. In $A_1$, the flux goes outward on all faces of the left tetrahedron,
but inward on the right tetrahedron. Arrows show a particular link realization. In
all other classes, colored faces have a flux opposite to that of $A_1$, which are realized
by flipping the colored links. The colors of the two faces $abc$ and $ab'c'$ are shown
in the two dotted rectangles. Flux configurations related by time reversal (i.e. flip-
ning all links on both tetrahedra) or tetrahedron symmetry (i.e. attaching the two
tetrahedra at a different vertex) are equivalent and not shown. Each tetrahedron
has an even number of flipped faces, entailing $\binom{4}{0} + \binom{4}{2} + \binom{4}{4} = 8$ flux configu-
rations, hence the pair has $8 \times 8 = 64$ configurations. The degeneracy of each class
arising from both TR and tetrahedral symmetry are noted in their labels. E.g.,
to determine the degeneracy of $C_2$, note that (i) there are $\binom{4}{2} = 6$ ways to fix the
right tetrahedron, (ii) for each right configuration, the left tetrahedron must have
one flipped face coplanar with one of the right flipped faces (multiplicity 2) and
the other non-coplanar with both flipped faces of the right (again 2), thus the total
multiplicity is $6 \times 2 \times 2 = 24$. For class $B$, one fixes the left tetrahedron first with
multiplicity $\binom{4}{2} = 6$. A multiplicity of 2 then arises from switching the left and
right, and a further multiplicity of 2 arises from TR, thus multiplicity of class $B$
is 24.
Figure 8.20: Hexagonal flux configuration on the pyrochlore lattice. Flipping colored links in every other layer of up-pointing tetrahedra can invert the pair of colored hexagonal fluxes without changing tetrahedral fluxes. Any pair of hexagonal fluxes can be changed in this way.

Thus for a given tetrahedral flux configuration, there are \((\binom{4}{0}) + (\binom{4}{2}) + (\binom{4}{4}) = 8\) hexagonal flux configurations. See Fig. 8.20 for a link realization.

### 8.7.3 MUC

In the MUC case, for each Kagome plane the hexagonal flux equals the product of the two triangular fluxes. The \(k\)-space Hamiltonian matrix is\(^{12}\)

\[
H(k) = i \sum_{\alpha, ss'} |s, \alpha\rangle \langle s', \alpha| \left[ J^\alpha_\Delta \Delta_{ss'} + J^\alpha_\triangledown \triangledown_{ss'} e^{+i k \cdot (r_s - r_{s'})} \right] + iK \sum_s \left[ |s, 1\rangle \langle s, 2| - h.c. \right]
\]

where \(\alpha = 1, 2\) labels the Majorana species, \(s, s' \in \{a, b, c, d\}\) label the vertices within each up-pointing tetrahedron, and \(\Delta_{ss'}\) and \(\triangledown_{ss'}\) denote signs of links (\(u_{ij}\)'s) along up- and down-pointing tetrahedra from sites \(s\) to \(s'\).

We may restrict ourselves to a two-parameter space,

\[
J_1 \equiv J^1_\Delta = J^1_\triangledown = \sin \theta \cos \phi \quad , \quad J_2 \equiv J^2_\Delta = J^2_\triangledown = \sin \theta \sin \phi \quad , \quad K = \cos \theta \quad ,
\]

\(^{12}\)Note that in the hopping term, the “+” sign is because the hopping displacement is from vertex \(s\) of the current tetrahedron to \(s'\) of the next tetrahedron, which is the reverse of the displacement from \(s\) to \(s'\) within the same tetrahedron.
Fig. 8.21 shows the ground state energy density of all pyrochlore flux classes at fixed $\phi = \frac{\pi}{4}$ ($J_1 = J_2 = \sin \theta / \sqrt{2}$).

For the MUC configurations considered here, we can identify trivial gapped, semimetallic, and metallic phases, cf. Fig. 8.22.

The gapped regions in all classes are located at $\theta \approx 0$, corresponding to the atomic limit ($K \gg J$, no hopping) and is therefore always trivial.

In the metallic phases, all three classes appear to have a normal nesting momentum $q = (\pi, \pi, \pi)$, cf. Figs. 8.23 and 8.24. However this maps their FSs back onto themselves due to the shape of the first BZ, hence there is no nesting instability.

### 8.8 Summary

In this chapter, we investigated Kitaev-type models in one, two, and three spatial dimensions. Kitaev-type models are models of interacting spins which can be mapped to free Majorana fermions hopping in static gauge fields. We reviewed this mapping using Kitaev’s original honeycomb model, and then restricted our-
Figure 8.22: Phase diagram of MUC pyrochlore. $\Delta E$ is the spectral gap at $E = 0$. Yellow line is $\Delta E$ for $\phi = 0.25\pi$ ($J_1 = J_2$). When $\Delta E = 0$, either the fourth and fifth bands touch at zero energy (semimetallic), or at least one band crosses zero (metallic). To distinguish the latter two cases we also plot $\Delta n$ which is the number of bands crossed by $E = 0$. A semimetallic phase has $\Delta n = 0$, whereas a metallic phase has $\Delta n > 0$. $\Delta n$ is always even due to the Majorana nature. From the $\Delta n$ plots, one concludes that $B$, $C_2$ and $C_3$ all have metallic phases. The “mini-gap” region in $A_2$ and $C_1$ near $\theta = 0.5\pi$ (non-hybridizing limit) is most probably a finite size artifact, because the magnitude of the gap is comparable with that of $C_3$ which in fact is in a metallic phase. Thus there is only one truly gapped phase for all classes in the atomic limit $J_{1,2} = 0$, which must be topologically trivial. 20 $\times$ 20 $\times$ 20 unit cells are used.
Figure 8.23: Fermi surface of class $B$ pyrochlore flux, with $\phi = 0.3\pi$ and different $\theta$, using $100 \times 100 \times 100$ $k$-space mesh. The $4^{th}$ and $5^{th}$ bands are related by $\varepsilon_4(k) = -\varepsilon_5(-k)$ due to the Majorana nature. Fermi surface of the $4^{th}$ band, $\varepsilon_4(k) = 0$, are shown in yellow, and that of the $5^{th}$ band in blue. Energy gradient (lower to higher) points from the wireframed side to the unframed side. We note that: (i) there appears to be a normal nesting momentum of $q = \pi(1,1,1)$, but this in fact maps the Fermi surface onto itself, thus there is no true nesting and the FS is stable; (ii) Although not shown here, at $\theta \sim 0.2\pi$, the two FSs shrink to points (or line nodes), causing the color switching between the $\theta = 0.19\pi$ and $0.21\pi$ shown here. Translucent gray bounding surface corresponds to the first BZ.
Figure 8.24: Fermi surface of classes $C_2$ and $C_3$ pyrochlore flux, with $\phi = 0.3\pi$ and $\theta = 0.1\pi$, using $50 \times 50 \times 50$ $k$-space mesh. The $4^{th}$ and $5^{th}$ bands are related by $\varepsilon_4(k) = -\varepsilon_5(-k)$ due to the Majorana nature. Fermi surface of the $4^{th}$ band, $\varepsilon_4(k) = 0$, are shown in yellow, and that of the $5^{th}$ band in blue. Energy gradient (lower to higher) points from the wireframed side to the unframed side. We note that: (i) there appears to be a normal nesting momentum of $q = \pi(1,1,1)$, but this in fact maps the Fermi surface onto itself, thus there is no true nesting and the FS is stable; (ii) Similar to class $B$, at $\theta \sim 0.2\pi$, the two FSs shrink to points (or line nodes), resulting a color switching when $\theta$ passes through $\sim 0.2\pi$. Translucent gray bounding surface corresponds to the first BZ.
selves to the fermion aspect of these models, which can be analyzed using band theory. We explained some features, properties, and quirkiness associated with the presence of Majorana fermions and gauge fields in simple 1D and 2D settings, such as the spectral symmetry in the presence of inversion symmetry, spectral redundancy arising from the Majorana nature, relation between energy and the flux configuration (Lieb's theorem), existence and stability of Fermi surfaces. With these possibilities in mind, we studied Kitaev models on two 3D lattices. The increased number of sites within a structural unit cell brings about more elementary plaquettes present in the model, making the $\mathbb{Z}_2$ flux configurations more fragmented, and we only considered flux configurations with high-enough translational symmetry, where flux unit cell consists of one, two or four structural unit cells. While such a strategy may preclude us from identifying the true ground state of the model if it resided in a $\mathbb{Z}_2$ flux sector that breaks the aforementioned translational invariance, we note that one can always stabilize the ground state of a “wrong” flux sector by adding plaquette terms such as Eq. 8.3 to the Hamiltonian, thus the quest for the true ground state is no longer a highly desired aim, and one instead turns to looking for states with more exotic quantum phases. Indeed we have identified (topologically trivial) gapped phase, semi-metallic phase with Weyl-node-like point Fermi surfaces, and metallic phase in these 3D models. The latter two phases, namely spin-semi-metal and spin-metal, would have been difficult to generate in 3D in the presence of interaction prior to the advent of Kitaev model.
Appendix A

Zigzag edge states of the Haldane model, and their application to graphene and boron nitride

A.1 Edge solution

From Eq. 2.34, we note that as $k_1 \to -k_1$,

$$v_1(-k_1) = v_2^*(k_1) \quad , \quad p_1(-k_1) = p_1(k_1) \quad (A.1)$$

Eq. 2.39 gives two equations (the ratio $r$ itself being yet undetermined) with two unknowns, $\lambda$ and $\varepsilon$, for each $k_1$ value. The second equality of Eq. 2.39 does not involve $\varepsilon$, and can be used to solve for $\lambda$,

$$p_2 v_2 \lambda^2 + (v_2 v_1^* - v_1 v_2^* + p_2) \lambda + p_2 v_1^* = 0, \quad (A.2)$$

This yields Eqs. 2.40 and 2.41. Note that sending $\lambda \leftrightarrow \lambda^{-1}$ and $v_1 \leftrightarrow v_2^*$ keeps Eq. A.2 invariant, thus together with Eq. A.1, one gets

$$d(k) = d(-k), \quad (A.3)$$

$$\lambda_+(k) = \lambda_-^{-1}(-k), \quad r_+(k) = r_-^{-1}(-k). \quad (A.4)$$

To solve for $\varepsilon$, we use a simple fact about ratios: if $r = x_1/y_1 = x_2/y_2$, then $r$ is preserved by arbitrary linear combination of numerators and denominators,
$r = (ax_1 + bx_2)/(ay_1 + by_2)$, except the unfortunate choice that makes $ax_1 + bx_2 = 0$. Applying to Eq. 2.39, we get

$$r = \frac{p_2 h_1 - p_2 \varepsilon - 2p_1 \text{Re} v_1}{p_2 h_2 - p_2 \varepsilon - 2p_1 \text{Re} v_2}$$  \hspace{1cm} (A.5)$$

where $\text{Re}$ indicates real part. This gives the two solutions of $\varepsilon$, Eq. 2.42, in terms of $r_\pm$ as solved in Eq. 2.41. Similar to Eq. A.3,

$$\varepsilon_+(m, k_1) = \varepsilon_-(-m, -k_1).$$  \hspace{1cm} (A.6)$$

The bulk spectrum has the same inversion property which is easily verified from Eq. 2.22.

In the text (Eq. 2.45) we mentioned that $r_\pm$ is real if the discriminant $\Delta$ is non-negative. The reality of $r_\pm$ (Eq. 2.45) in turn has the following implication: Using $r = r^*$ and Eq. 2.39,

$$r = \frac{v_1^* + \lambda p_2}{v_2^*} = r^* = \frac{\lambda^* v_1^*}{\lambda^* v_2^* + p_2} \implies r = -|\lambda|^2,$$  \hspace{1cm} (A.7)$$

which can also be checked explicitly using Eq. 2.41. Notice that this applies to the singular case, Eq. 2.44, as well.

That we can get Eq. A.7 is actually fortunate, for otherwise there will be no twisted boundary consistent with the $\text{Ansatz}$. We will come back to this point later when deriving the eigenstates.

### A.2 Edge crossing point

To derive the edge crossing point, we note that according to Eq. 2.41, the ratio $r$ depends on the branch: in general $r_+ \neq r_-$. But Eq. A.5 implies that at the edge crossing point(s), $r_+ = r_-$. It would seem the latter is satisfied only when the discriminant $\Delta = 0$, but one can easily verify from Fig. 2.5 that these do not correspond to the edge crossing points. In fact, there is a range of parameters in the topological phase where $\Delta$ is always positive, e.g., $t_1 = t_2 = t_3 = 0.3$, $m = \phi/\pi = 0.5$. Recall that in deriving Eq. A.5, we used linear combinations of denominators and numerators of the three ratios in Eq. 2.39, the validity of which
requires these combinations to be non-singular (cf. discussion leading to Eq.A.5). Thus the only way for Eqs. 2.41 and A.5 to be consistent—the former implying $r_+ \neq r_-$, the latter implying otherwise—is for the linear combinations of both the denominators and the numerators to be singular, viz.,

$$p_2(h_1 - \varepsilon) - 2p_1 \text{Re} v_1 = p_2(h_2 - \varepsilon) - 2p_1 \text{Re} v_2 = 0.$$  \hspace{1cm} (A.8)

This yields the edge-crossing condition Eq. 2.46.

\section*{A.3 Eigenstates}

Eqs. 2.37 and 2.38 can be cast into the same Schrödinger equation,

$$a \psi_{n+1} + b \psi_n + c \psi_{n-1} = 0,$$  \hspace{1cm} (A.9)

where $a$, $b$ and $c$ can either be the set of numerators or denominators in Eq. 2.39, e.g., $a = v_1$, $b = h_1 + \lambda p_1 - \varepsilon$, $c = v_1^* + \lambda p_2$. Note that these are already known once the edge solutions are obtained. Eq. A.9 is equivalent to

$$\psi_{n+1} - x_1 \psi_n = x_2 (\psi_n - x_1 \psi_{n-1}),$$  \hspace{1cm} (A.10)

$$x_1 + x_2 = -\frac{b}{a}, \quad x_1 x_2 = \frac{c}{a},$$  \hspace{1cm} (A.11)

i.e., $x_1$ and $x_2$ are solutions to

$$ax^2 + bx + c = 0.$$  \hspace{1cm} (A.12)

Denote $\phi_n = \psi_n - x_1 \psi_{n-1}$, then by eqn A.10, $\phi_n = x_2^{n-1} \phi_1$, and

$$\psi_n = \phi_n + x_1 \psi_{n-1} = \phi_n + x_1 \phi_{n-1} + x_1^2 \psi_{n-2}$$

$$= \ldots$$

$$= \phi_n + x_1 \phi_{n-1} + x_1^2 \phi_{n-2} + \ldots$$

$$+ x_1^{n-2} \phi_2 + x_1^{n-1} \underbrace{\left( \phi_1 + x_1 \psi_0 \right)}_{\psi_1}.$$  \hspace{1cm} (A.13)

1. If $x_1 \neq x_2$, the geometric series can be summed,

$$\psi_n = \frac{x_1^n - x_2^n}{x_1 - x_2} \phi_1 + x_1^n \psi_0 = f_n \psi_1 - x_1 x_2 f_{n-1} \psi_0,$$  \hspace{1cm} (A.14)

$$f_n \equiv \frac{x_1^n - x_2^n}{x_1 - x_2}.$$  \hspace{1cm} (A.15)
2. If \( x_1 = x_2 = x \), then

\[
\psi_n = nx^{n-1}\psi_1 - (n - 1)x^n\psi_0 ,
\]  

(A.16)

which is the same as applying L'Hospital's rule on the previous case.

3. If \( ac = 0 \), one of the \( x_i \), say \( x_2 \), is 0, then

\[
\psi_n = x_1^{n-1}\psi_1 .
\]  

(A.17)

Thus in all cases we may proceed with Eq. A.14. This yields

\[
\psi_{N+1} = f_{N+1}\psi_1 - x_1x_2f_N\psi_0 ,
\]  

(A.18)

\[
\psi_N = f_N\psi_1 - x_1x_2f_{N-1}\psi_0 .
\]  

(A.19)

We now come to the issue of boundary conditions. Eq. 2.35 implies

\[
v^\dagger \begin{pmatrix}
\psi_0 \\
\lambda\psi_0
\end{pmatrix} = \rho v^\dagger U^\dagger \begin{pmatrix}
\psi_N \\
\lambda\psi_N
\end{pmatrix} ,
\]  

(A.20)

\[
v \begin{pmatrix}
\psi_{N+1} \\
\lambda\psi_{N+1}
\end{pmatrix} = \rho U v \begin{pmatrix}
\psi_1 \\
\lambda\psi_1
\end{pmatrix} .
\]  

(A.21)

From Eq. 2.27,

\[
v^{-1} = \frac{1}{v_1v_2} \begin{pmatrix}
v_2 & 0 \\
-p_2 & v_1
\end{pmatrix}
\]  

(A.22)

which exists if \( v_1v_2 \neq 0 \). The case \( v_1v_2 = 0 \) can be analyzed by Taylor expanding in the vanishing \( v_i \) similar to the discussion leading to Eq. 2.44. The \( v^\dagger \) can be dropped from Eq. A.20. For Eq. A.21, one gets from Eq. 2.39 that

\[
v \begin{pmatrix}
1 \\
\lambda
\end{pmatrix} = \frac{v_1}{r} \begin{pmatrix}
r \\
\lambda
\end{pmatrix} ,
\]  

(A.23)

thus the two boundary conditions Eqs. A.20 and A.21 become

\[
\psi_0 \begin{pmatrix}
1 \\
\lambda
\end{pmatrix} = \rho \psi_N U^\dagger \begin{pmatrix}
1 \\
\lambda
\end{pmatrix} , \quad \psi_{N+1} \begin{pmatrix}
r \\
\lambda
\end{pmatrix} = \rho \psi_1 U \begin{pmatrix}
r \\
\lambda
\end{pmatrix} .
\]  

(A.24)
This implies both \((1_\lambda)\) and \((r_\lambda)\) are eigenstates of \(U\),

\[
U^\dagger \begin{pmatrix} 1 \\ \lambda \end{pmatrix} \equiv e^{-i\theta_1} \begin{pmatrix} 1 \\ \lambda \end{pmatrix}, \quad U \begin{pmatrix} r \\ \lambda \end{pmatrix} \equiv e^{i\theta_2} \begin{pmatrix} r \\ \lambda \end{pmatrix}.
\] (A.25)

Thus, if \(U \neq e^{i\theta} I\), then \((1_\lambda)\) and \((r_\lambda)\) are either equivalent (if \(\theta_1 = \theta_2\)), indicating \(1 = r\), or orthogonal (if \(\theta_1 \neq \theta_2\)), indicating \(r = -|\lambda|^2\). The latter is nothing but Eq. A.7. The boundary conditions then reduce to

\[
\psi_0 = \rho e^{-i\theta_1} \psi_N, \quad \psi_{N+1} = \rho e^{i\theta_2} \psi_1.
\] (A.26)

Substituting these in Eq. A.14 yields the following condition,

\[
\frac{x_1 x_2 f_N}{f_{N+1}} e^{i(\theta_2 - \theta_1)} \rho^2 \equiv A
\] 
\[
+ \left[ \frac{x_1 x_2}{f_{N+1}} e^{-i\theta_1} + \frac{1}{f_{N+1}} e^{i\theta_2} \right] \rho - 1 = 0
\] (A.27)

where one needs to tune \(\theta_1\) and \(\theta_2\) such that at least one solution of \(\rho\) is real. A sufficient condition is for both \(A\) and \(B\) to be real. Here we make \(A > 0\) so that the solutions of \(\rho\) are always real,

\[
\theta_2 - \theta_1 = \text{arg} \left( \frac{f_{N+1}}{x_1 x_2 f_{N-1}} \right), \quad A = \left| \frac{x_1 x_2 f_{N-1}}{f_{N+1}} \right|,
\] (A.28)

then

\[
B = B_1 e^{-i\theta_1} + B_2 e^{i\delta} e^{i\theta_1}
\] (A.29)

thus for \(B\) to be real,

\[
0 = B - B^* = (B_1 - B_2^* e^{-i\delta}) e^{-i\theta_1} - (B_1 - B_2^* e^{-i\delta})^* e^{i\theta_1}
\] (A.30)

\[
\implies \theta_1 = \text{arg} \left( B_1 - B_2^* e^{-i\delta} \right) + \left\{ \begin{array}{c} 0 \\ \pi \end{array} \right\}
\] (A.31)

where the freedom \(\left\{ \begin{array}{c} 0 \\ \pi \end{array} \right\}\) can be used to switch the sign of \(B\) (which has been made real). Then we have

\[
\rho_\pm = -\frac{|B|}{2A} \pm \frac{\sqrt{|B|^2 + 4A}}{2A}
\] (A.32)
Since $|B|$ and $A$ are both $\geq 0$, the $\rho$ with smaller magnitude (which has a better chance of $\to 0$ to represent an open boundary) is always $\rho_+$. Notice that each of $\rho_\pm$ still depends on which branch of the edge solution we have picked in calculating $x_1$ and $x_2$.

In particular, one can show that $\rho = 1$ and $|x_i| = 1$ occur simultaneously, the former means periodic boundary condition, while the latter means the Ansatz solution is a bulk solution, so this makes sense. To see this, assume $\rho = 1$ and $x_2 = e^{i\phi}$, then Eq. A.27 becomes

$$[e^{iN\phi} e^{-i\theta_1} - 1] x_{N+1} + e^{i\phi} [e^{i(\theta_2 - \theta_1)} - e^{iN\phi} e^{-i\theta_1}] x_N^N$$

$$+ [e^{i\theta_2} - e^{iN\phi} e^{i(\theta_2 - \theta_1)}] - [e^{i\phi} e^{i\theta_2} - e^{i(N+1)\phi}] = 0. \quad (A.33)$$

One then gets $e^{i\theta_1} = e^{i\theta_2} = e^{iN\phi}$ by requiring all square brackets to be zero, i.e., Eq. A.27 is indeed consistent.

**A.4 Graphene and boron nitride**

For graphene, $K^{(1)} = K^{(2)} = 0$, so Eqs. 2.37 and 2.38 become

$$\lambda R \psi_A = \varepsilon \psi_A, \quad R^\dagger \psi_A = \lambda \varepsilon \psi_A, \quad (A.34)$$

$$R = \begin{pmatrix} p_1 & z p_2 \\ p_2 & p_1 \\ \vdots & \vdots \\ p_2 & p_1 \end{pmatrix}, \quad (A.35)$$

$$p_1 = -2 \cos \frac{k}{2}, \quad p_2 = -1 \quad (A.36)$$

where $z$ controls the boundary condition.

If $\lambda \neq 0$ and $1/\lambda \neq 0$, i.e., both $A$ and $B$ sites have charge density, then $R \psi_A = \varepsilon \psi_A/\lambda$ gives

$$p_2 \psi_n + p_1 \psi_{n+1} = \frac{\varepsilon}{\lambda} \psi_{n+1}, \quad (A.37)$$

$$\psi_0 = z \psi_N, \quad n = 0, 1, 2, \ldots, N \quad (A.38)$$
Thus
\[ \psi_{n+1} = \left( \frac{p_2}{\frac{\varepsilon}{\lambda} - p_1} \right)^{n+1} \psi_0 , \quad z = \left( \frac{p_2}{\frac{\varepsilon}{\lambda} - p_1} \right)^{-N} , \]  
(A.39)

whereas \( R^\dagger \psi_A = \lambda \varepsilon \psi_A \) will give
\[ p_1 \psi_n + p_2 \psi_{n+1} = \lambda \varepsilon \psi_n , \]  
(A.40)
\[ \psi_{N+1} = z \psi_N , \quad n = 1, 2, \ldots, N, N+1 \]  
(A.41)
yielding
\[ \psi_{n+1} = \left( \frac{\lambda \varepsilon - p_1}{p_2} \right)^{n} \psi_1 , \quad z = \left( \frac{\lambda \varepsilon - p_1}{p_2} \right)^N , \]  
(A.42)

Equating the expressions for \( z \) gives
\[ \lambda = \pm 1 \implies z = \left( \frac{\tilde{\varepsilon} - p_1}{p_2} \right)^N \]  
(A.43)

where \( \tilde{\varepsilon} \equiv \lambda \varepsilon = \varepsilon / \lambda \). Equating the two recursions then yields
\[ \tilde{\varepsilon} = p_1 \pm p_2 = -2 \cos \frac{k}{2} + 1 , \quad z = (\pm 1)^N = \pm 1. \]  
(A.44)

Since \( |z| = 1 \), these states correspond to bulk states with periodic or antiperiodic boundary conditions.

If \( \lambda = 0 \), then
\[ \begin{pmatrix} \begin{bmatrix} R \\ R^\dagger \end{bmatrix} \end{pmatrix} \begin{pmatrix} \psi_A \\ 0 \end{pmatrix} = \begin{pmatrix} \varepsilon \psi_A \\ 0 \end{pmatrix} \]  
(A.45)
\[ \implies \varepsilon = 0 , \quad \psi_{n+1} = -\frac{p_1}{p_2} \psi_n \]  
(A.46)

and
\[ z = -\left( \frac{p_1}{p_2} \right)^N = -(2 \cos \frac{k}{2})^N \]  
(A.47)
\[ \implies \begin{cases} |z| \ll 1 & k \in (\frac{2\pi}{3}, \frac{4\pi}{3}) \\ |z| \gg 1 & k \in (-\frac{2\pi}{3}, \frac{2\pi}{3}) \end{cases} , \]  
(A.48)

thus it is an edge solution localized on the \( A \) sites at the \( y = 1 \) edge if \( |k| > 2\pi/3 \) (\( i.e. \), connecting two Dirac points), with zero energy.
Similarly, if $1/\lambda = 0$, then

$$\begin{pmatrix} R \\ R^\dagger \end{pmatrix} \begin{pmatrix} 0 \\ \psi_B \end{pmatrix} = \varepsilon \begin{pmatrix} 0 \\ \psi_B \end{pmatrix}$$

(A.49)

$$\implies \varepsilon = 0, \quad \psi_{n+1} = -\frac{p_2}{p_1} \psi_n$$

(A.50)

and

$$z = -\left(\frac{p_1}{p_2}\right)^N = -(2 \cos \frac{k}{2})^N$$

(A.51)

$$\implies \begin{cases} |z| \ll 1 & k \in (\frac{2\pi}{3}, \frac{4\pi}{3}) \\ |z| \gg 1 & k \in (-\frac{2\pi}{3}, \frac{2\pi}{3}) \end{cases}$$

(A.52)

thus it is an edge solution localized on the $B$ sites at the $y = N$ edge if $|k| > 2\pi/3$ with zero energy.

For boron nitride (BN), the $A$ sublattice is nitrogen and $B$ is boron. If $\lambda \neq 0$, $1/\lambda \neq 0$, then $\lambda R\psi_A = (\varepsilon - m)\psi_A$ yields

$$\psi_{n+1} = \frac{p_2}{\varepsilon_- - p_1} \psi_n, \quad z = \left(\frac{\varepsilon_- - p_1}{p_2}\right)^N$$

(A.53)

where $\varepsilon_- \equiv \frac{\varepsilon - m}{\lambda}$. Similarly $R^\dagger \psi_A = \lambda(\varepsilon + m)\psi_A$ gives

$$\psi_{n+1} = \frac{\varepsilon_+ - p_1}{p_2}, \quad z = \left(\frac{\varepsilon_+ - p_1}{p_2}\right)^N$$

(A.54)

where $\varepsilon_+ \equiv \lambda(\varepsilon + m)$. Equating expressions for $z$ gives

$$\frac{\varepsilon_- - p_1}{p_2} = \frac{\varepsilon_+ - p_1}{p_2}$$

(A.55)

while equating the two $\psi$ recursions gives

$$\frac{\varepsilon_- - p_1}{p_2} = \frac{p_2}{\varepsilon_+ - p_1}$$

(A.56)

thus

$$\frac{\varepsilon_- - p_1}{p_2} = \frac{\varepsilon_+ - p_1}{p_2} = \pm 1$$

(A.57)

$$\implies \varepsilon = \pm \sqrt{m^2 + (p_1 \pm p_2)^2}, \quad |z| = 1.$$  \hspace{1cm} (A.58)

When $|z| = 1$, these are bulk states.

If $\lambda = 0$, then the edge solution is the same as the corresponding graphene case with $\varepsilon = m$ and the edge state is purely on $A$ sites. If $1/\lambda = 0$, then $\varepsilon = -m$ and the edge state is purely on $B$ sites.
Appendix B

Incommensurate edge spectrum in the Hofstadter model

Here we first briefly review the edge spectrum with commensurate $N_y$ and $q$ as studied in ref. [27], and then extend its argument to the incommensurate case.

B.1 Review of commensurate edge spectrum

The Schrödinger equation corresponding to the matrix equation $H(k_x)_{yy'} \psi_{y'} = \varepsilon \psi_y$ is

$$-\psi_{y-1} - \psi_{y+1} - 2 \cos(k_x + y\phi) \psi_y = \varepsilon \psi_y .$$

(B.1)

Cast into transfer matrix form, we have

$$\begin{pmatrix} \psi_{y+1} \\ \psi_y \end{pmatrix} = M_y \begin{pmatrix} \psi_y \\ \psi_{y-1} \end{pmatrix},$$

(B.2)

$$M_y = \begin{pmatrix} -\varepsilon - 2 \cos(k_x + y\phi) & -1 \\ 1 & 0 \end{pmatrix}$$

(B.3)

Notice that $M_y$ depends on $\varepsilon$. The following boundary condition is required for Eq. B.2 to also cover the cases of $y = 1$ and $N_y$,

$$\psi_0 = \psi_{N_y+1} = 0 .$$

(B.4)
Then

$$\begin{pmatrix} \psi_{N_y+2} \\ \psi_{N_y+1} \end{pmatrix} = \mathcal{M}_{N_y+1} \begin{pmatrix} \psi_1 \\ \psi_0 \end{pmatrix} \quad (B.5)$$

and Eq. B.4 implies $M_{N_y+1}$ is a triangular matrix,

$$[\mathcal{M}_{N_y+1}]_{21} = 0 \quad (B.7)$$

The spectrum $\{\varepsilon\}$ consists of all energies satisfying Eq. B.7.

Notice that $M_{y+q} = M_y$, so when $N_y + 1 = qL$ with integer $L$ ("commensurate"),

$$\mathcal{M}_{N_y+1} = Q^L, \quad Q \equiv M_q \quad (B.8)$$

Now, products of upper-triangular matrices are still upper-triangular, so Eq. B.7 is satisfied if $Q$ is upper-triangular,

$$Q_{21} = 0 \quad (B.9)$$

It is then easy to verify that

$$\psi_{\ell q+1} = [Q_{11}]^\ell \psi_1, \quad \psi_{\ell q} = 0 \quad (B.10)$$

where $\ell = 1, 2, \ldots, L$, hence the solution is an edge state exponentially localized at $y = N_y$ if $|Q_{11}| > 1$, and at $y = 1$ if $|Q_{11}| < 1$.

The edge spectrum $\{\varepsilon\}$ satisfying the condition $Q_{21}(\varepsilon) = 0$ is the same as the full spectrum of a $(q - 1) \times (q - 1)$ system, so numerically the edge spectrum of $H(k_x, N_y = Lq - 1, z = 0)$ can be solved by diagonalizing its upper-left $(q - 1) \times (q - 1)$ submatrix.

Note that Eq. B.10 implies the edge states, with $\psi_0$ included, have a direct
product form

\[ |\psi\rangle = \left( \begin{array}{c} Q_{11}^0 \\ Q_{11}^1 \\ Q_{11}^2 \\ \vdots \\ Q_{11}^{L-1} \end{array} \right) \otimes \left( \begin{array}{c} \psi_0 \\ \psi_1 \\ \psi_2 \\ \vdots \\ \psi_{q-1} \end{array} \right), \quad (B.11) \]

i.e., \( \psi_{\ell q+m} = Q_{11}^\ell \psi_m \) with \( \ell = 0, 1, \ldots, L - 1 \) and \( m = 0, 1, \ldots, q - 1 \). The \( N \)-component magnetic cell part dictates the real-space behavior. In this case it is exponentially localized at either end. The \( q \)-component internal part is obtained by prepending \( \psi_0 = 0 \) to the solutions of the \((q-1) \times (q-1)\) upper-left block of \( H \). This is by no means a general form of edge states, but we do also notice a similar decomposition in the zigzag edge modes of the Haldane model, see Chapter 2. Note also that all Bloch states have such a decomposition, \( |\Psi(k,n)\rangle = |k\rangle \otimes |\psi_n(k)\rangle \) where \( \langle y|k\rangle = e^{i k y}/\sqrt{N} \) is the Bloch phase and \( |\psi_n(k)\rangle \) is the \( n \)th band eigenstate of the Fourier transformed \( q \times q \) Hamiltonian. One may then say that \(-i \log(Q_{11})\) is the imaginary Bloch vector, and which of the UHP or LHP it resides in tells the localization of the edge states.

### B.2 Incommensurate edge spectrum

In the thermodynamic limit where \( N_y \to \infty \), one can extend the commensurate argument to incommensurate cases, \( N_y + 1 = Lq + m \), with \( m = 0, 1, 2, \ldots, q - 1 \).

First, we note two properties of the transfer matrix,

\[ \det(M_y) = 1 \quad (B.12) \]
\[ M_{y+m}(k_x, \epsilon) = M_y(k_x + m\phi, \epsilon) \quad (B.13) \]

both are straightforward from the definition. Eq. B.13 expresses the same \( k_x \) covariance as Eq. 5.20. The \((Lq + m)\)-step transfer matrix can then be divided in
two ways,

\[ \mathcal{M}_{Lq+m}(k_x) = \mathcal{M}_m(k_x) Q^L(k_x) \]

\[ = Q^L(k_x + m\phi) \mathcal{M}_m(k_x) \]  

(B.14)

(B.15)

If \( Q(k_x) \) satisfies the commensurate edge condition Eq. B.9, then

\[ Q^L(k_x) = \begin{pmatrix} [Q_{11}]^L & x \\ 0 & [Q_{22}]^L \end{pmatrix}_{k_x}, \]

(B.16)

where \( x \) is some number of no interest. We then have

\[ \mathcal{M}_{Lq+m}(k_x) = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}_{k_x} \begin{pmatrix} [Q_{11}]^L & x \\ 0 & [Q_{22}]^L \end{pmatrix}_{k_x}, \]

(B.17)

hence

\[ [\mathcal{M}_{Lq+m}]_{21}(k_x) = A_{21} [Q_{11}]^L (k_x) \]

\[ = A_{21} [Q_{22}]^{-L} (k_x) \]

(B.18)

where the second equality follows from \( \text{det} \ Q = 1 \), a consequence of Eq. B.12.

Similarly, if \( Q(k_x + m\phi) \) satisfies the commensurate edge condition Eq. B.9, we have instead

\[ \mathcal{M}_{Lq+m}(k_x) = \begin{pmatrix} [Q_{11}]^L & x \\ 0 & [Q_{22}]^L \end{pmatrix}_{k_x+m\phi} \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \]

(B.19)

thus

\[ [\mathcal{M}_{Lq+m}]_{21}(k_x) = [Q_{22}]^L (k_x + m\phi) A_{21} \]

\[ = [Q_{11}]^{-L} (k_x + m\phi) A_{21} \]

(B.20)

We can then conclude that in the limit \( L \to \infty \),

1. If \((k_x, \varepsilon)\) is a solution of the commensurate case \( N_y + 1 = Lq \) at the lower edge, \([Q_{11}]^L (k_x) \to 0\), then by Eq. B.18, it is also a solution of incommensurate \( N_y + 1 = Lq + m \). For \( y \in [1, Lq - 1] \), the wavefunction \( \psi_y \) coincides with that of the commensurate case, and in the upper tail where \( y = Lq + m \), \( \psi_y = A_{11} [Q_{11}]^L \psi_1 \to 0 \), thus it is also at the lower edge.
2. If \((k_x, \varepsilon)\) is a solution of \(N_y + 1 = qL\) at the upper edge, \([Q_{11}]^{-L}(k_x) \rightarrow 0\), then by Eq. B.20, \((k_x - m\phi, \varepsilon)\) will be a solution of incommensurate \(N_y + 1 = Lq + m\). It is also at the upper edge because \(\psi_y \rightarrow 0\) in the lower tail.

Thus for \(N_y \rightarrow \infty\), an increment of \(N_y\) by 1 leaves the lower edge spectrum unchanged, while shifting the upper edge spectrum in \(k_x\) by \(-2\pi p/q\).
Appendix C

Triangular lattice Hofstadter model

The Hofstadter model on a triangular lattice can be obtained by adding in each square plaquette a diagonal bond along the $\hat{x} - \hat{y}$ direction with half-odd-integer vector potential (in units of $\phi = 2\pi p/q$), such that the flux per triangle is $\phi/2$. Instead of Eq. 5.5, the Hamiltonian matrix is now

$$H(k_x, N_y, z) = -\begin{pmatrix} c_1 & v_1 & 0 & \cdots & z^*v_{N_y-1}^* \\ v_1^* & c_2 & v_2 & 0 & \vdots \\ 0 & v_2^* & \ddots & \vdots \\ \vdots & & & v_{N_y-1} \\ zv_{N_y} & 0 & \cdots & v_{N_y-1}^* & c_{N_y} \end{pmatrix}$$

where

$$c_y = 2 \cos(k_x + y\phi) \quad , \quad v_y = 1 + t' e^{-i(k_x + y\phi + \frac{1}{2}\phi)} \quad , \quad (C.2)$$

with $t' = 1$ for triangular lattice, and 0 for square lattice. The discussion of edge spectrum in Appendix B remains essentially the same, except the transfer matrix, Eq. B.3, now becomes

$$M_y = \begin{pmatrix} -\varepsilon + c_y & -v_{y-1}^* \\ v_y & v_y \\ 1 & 0 \end{pmatrix}$$

(C.3)
Figure C.1: Energy and entanglement spectrum of the Hofstadter model on a triangular lattice with cylindrical boundary condition. Parameters used are $p/q = 5/8$, $N_y = 32$ and $E_F = 1$. In the energy spectrum (top panel), black dots represent bulk levels, red lines represent edge modes localized along the lower edge ($y = 1$), and blue lines represent those localized along the upper edge ($y = N_y = 32$). Vertical gray line indicates the $k_x$ value at which $E_F$ intersects the lower edge state. The entanglement occupancies (bottom panel) are computed for the lower half of the cylinder ($1 \leq y \leq 16$), color and symbol-coded according to $a$. The sudden color change happens when the lower edge mode crosses $E_F$. 
Figure C.2: Band projectors for the $p/q = 5/8$ Hofstadter model on a triangular lattice, $N_y = 32$, with periodic boundary conditions in $y$ and $k_x = 2\pi k/q$. $B_j$ and $G_\nu$ are projectors for the $j$th band and $\nu$ lowest bands, respectively, see also Fig. 5.3. Here we only plot the $24 \times 24$ submatrix belonging to the first three magnetic unit cells. Since the Hamiltonian Eq. C.1 can no longer be made purely real, the projectors are in general complex, so we only plot the real part of their matrix elements. Their signs are represented by color, red for positive and blue for negative, and their magnitudes represented by intensity. The individual bands $B_1$ and $B_2$ are better localized at $\kappa = 1$ where the gap between them is maximal, while their sum, $G_2$, is better localized at $\kappa = 1.5$ where the gap between them is minimal. Note that at both $\kappa$, the off-diagonal subblocks of both $B_j$ within each $8 \times 8$ block tend to cancel (they have different colors). The adiabatic evolution of the wave packets are obvious: for example, at $\kappa = 1.5$, the wave packets of $G_2$, in each unit cell, are at $y = 3$ and 6 (mod $q$. Same below). At $\kappa = 2$, the wave packet at $y = 3$ is in the progress of moving toward $y = 1$ while the one at $y = 6$ is “frozen”. At $\kappa = 2.5$, the first wave packet arrives at $y = 1$. In the next $\kappa \to \kappa + 1$ sub-cycle, the first wave packet will be frozen and the second one will migrate in the diagonal line by $-2$ (negative of the total Chern number).
In the pathological case where certain $v_y = 0$, the open edge Hamiltonian (i.e., $z = 0$) reduces to two blocks, $\{1, \ldots, \tilde{y}\}$ and $\{\tilde{y} + 1, \ldots, N_y\}$, each of which can be individually solved; alternatively one can shift $t$ slightly away from 1. Note that while $\text{det} M_y$ is no longer one, the $q$-step transfer matrix $Q$ still has unimodular determinant,

$$|\text{det} Q| = \left| \frac{v_0^* v_1^* \cdots v_{q-1}^*}{v_1 v_2 \cdots v_q} \right| = 1,$$

where we used $v_0 = v_q$. Consequently,

$$|Q_{22}| = |Q_{11}|^{-1},$$

and Eqs. B.18 and B.20 hold up to a phase. The conclusion thus remains unchanged that lower edge states are unchanged while upper edge states shift in $k_x$ with incommensurate $N_y$.

In Fig. C.1, we plot the cylindrical boundary energy spectrum, and its entanglement spectrum with $\nu = 5$ filled bulk bands for $p/q = 5/8$ on the triangular lattice. The Chern numbers of individual bands are either $C = -3$ or 5, which are equivalent modulo $q = 8$, and the lowest band has $C = -3$. Both are in agreement with what we observed in the square lattice case, namely, all band Chern numbers are equivalent modulo $q$, and that the lowest band Chern number is the one with smaller magnitude. The number of edge spectral flows in each gap is the total band Chern number below the gap, and the entanglement spectral flow mimics the behavior of edge spectral flow, and has an index discontinuity at $k_x$ where Fermi energy intersects the lower edge mode. Note that for $\nu = 5$ filled bands, the total Chern number is 1, as reflected in the number of edge and entanglement spectral flows. This agrees with our observation that the total Chern number of $p$ filled bands is one, see discussion following Eq. 5.25 in the text.

As in the square lattice case, the band projectors and their sums also flow under adiabatic $k_x$ pumping, and the number of wave packets crossing any given boundary during one cycle of the pumping reflects the Chern number of the projectors. Fig. C.2 shows the flow of the lowest two band projectors, $B_1$ and $B_2$, and their sum $G_2$, at $k = \kappa \cdot 2\pi/q$ with $\kappa = 1$ and 1.5. Both $B_j$ and $G_\nu$ have better
localization at either integer or half-odd-integer $\kappa$ where its gap from neighboring bands are maximal.

We thus conclude that the observations as detailed in the text using square lattice are robust and insensitive to the underlying lattice used.
Appendix D

Green’s functions of discrete and continuum Weyl Semimetal models

D.1 Spectrum and Green’s function of Eq. 7.2

Here we derive the spectrum and Green’s function of a generic Hamiltonian

\[ H(k) = \xi(k) + \tilde{H}(k) \]  

(D.1)

where

\[ \tilde{H} \equiv \sum_{a=1}^{5} d_a(k) \Gamma^a + h \Gamma^{\mu\nu} \]  

(D.2)

Note that the index \( a \) goes from 1 to 5, thus \( d_4(k) \) would be \( m(k) \) in Eq. 7.2. The symmetry breaking term is either \( \tilde{\Gamma} = \Gamma^{\mu\nu} \), in which case its strength is \( \eta = h \), or \( \tilde{\Gamma} = \Gamma^a \) in which case \( \eta = d_a \). In the following, \( k \) dependence will be suppressed.

From Eq. D.2, it is easy to verify that

\[ \tilde{H}^2 = d^2 + h^2 + 2h \Gamma^{\mu\nu} d_{\perp} \cdot \Gamma_{\perp} \]  

(D.3)

where \( d^2 \equiv \sum_{a=1}^{5} d_a^2 \) and \( d_{\perp} \) denotes the three components “perpendicular” to the \( \mu\nu \) “plane”, viz., \( d_{\perp} \cdot \Gamma_{\perp} = \sum_{a=1}^{5} d_a \Gamma^a - d_{\mu} \Gamma^\mu - d_{\nu} \Gamma^\nu \). The “parallel” components
vanish in the cross term due to their anticommutation with $\Gamma^{\mu\nu}$. Moving the scalars to the left hand side and squaring again yields

$$ (\tilde{H}^2 - d^2 - h^2)^2 = 4h^2d_{\perp}^2, \quad (D.4) $$

where we have used $[\Gamma_{\perp}, \Gamma^{\mu\nu}] = 0$. Here $d_{\perp}$ is the magnitude of $d_{\perp}$. Replacing $\tilde{H}$ with its eigenvalues $\tilde{E} = E - \xi$ gives the spectrum of $H$,

$$ E = \xi \pm \sqrt{d^2 + h^2 \pm 2hd_{\perp}}. \quad (D.5) $$

The Green’s function of Eq. D.1 is (denoting $\tilde{\omega} = \omega - \xi$)

$$ G(\omega) = \frac{1}{\omega - \tilde{H}} = \frac{1}{\tilde{\omega}^2 - \tilde{H}^2} = \frac{\tilde{\omega}^2 + \tilde{H}}{\tilde{\omega}^2 - \tilde{H}^2} = \frac{(\tilde{\omega} + \tilde{H})(\tilde{\omega} - 2d^2 - 2h^2 + \tilde{H}^2)}{(\tilde{\omega}^2 - d^2 - h^2)^2 - (\tilde{H}^2 - d^2 - h^2)^2} \equiv \frac{M}{D}. \quad (D.6) $$

Using Eq. D.4, the denominator $D$ is a number,

$$ D = (\tilde{\omega}^2 - d^2 - h^2)^2 - 4h^2d_{\perp}^2 \quad (D.8) $$

which is nothing but $\prod_i (\omega - E_i)$ with $E_i$ given by Eq. D.5. The numerator in powers of $\tilde{H}$ is

$$ M = \tilde{\omega}(\tilde{\omega}^2 - 2d^2 - 2h^2) + (\tilde{\omega}^2 - 2d^2 - 2h^2)\tilde{H} + \tilde{\omega}\tilde{H}^2 + \tilde{H}^3, \quad (D.9) $$

in which $\tilde{H}^2$ is already given by Eq. D.3, and

$$ \tilde{H}^3 = \tilde{H}\tilde{H}^2 = (d^2 + h^2)\tilde{H} + 2h d_{\perp} \cdot \Gamma_{\perp} \sum_a d_a \Gamma^a \Gamma^{\mu\nu} + 2h^2d_{\perp} \cdot \Gamma_{\perp}. \quad (D.10) $$

Rewriting $\sum_a d_a \Gamma^a = d_{\perp} \cdot \Gamma_{\perp} + d_{\mu} \Gamma^\mu + d_{\nu} \Gamma^\nu$, and using $\Gamma^{\mu(\nu)} \Gamma^{\mu\nu} = +(-)i\Gamma^{\nu(\mu)}$, we have

$$ \tilde{H}^3 = (d^2 + h^2)\tilde{H} + 2h d_{\perp} \cdot \Gamma^{\mu\nu} + 2h(h + id_{\mu} \Gamma^\nu - id_{\nu} \Gamma^\mu) d_{\perp} \cdot \Gamma_{\perp}. \quad (D.11) $$

Substituting Eqs. D.2, D.3 and D.11 in Eq. D.9 gives

$$ M = \tilde{\omega}(\tilde{\omega}^2 - d^2 - h^2) + (\tilde{\omega}^2 - d^2 - h^2)\tilde{H} + 2h d_{\perp} \cdot \Gamma^{\mu\nu} \\
+ 2h \left[ \tilde{\omega} \Gamma^{\mu\nu} + h + id_{\mu} \Gamma^\nu - id_{\nu} \Gamma^\mu \right] d_{\perp} \cdot \Gamma_{\perp}. \quad (D.12) $$
Eqs. D.8 and D.12 can now be used to obtain the local Green’s functions. Note that since \( d_i(-k) = -d_i(k) \) for \( i = 1, 2, 3 \) in the Hamiltonian of Eq. 7.2, many terms in Eq. D.12 will vanish upon \( k \)-space averaging.

(1) If \( \Gamma = \Gamma^\mu, \mu \neq 4 \) (see Sec. 7.5.2), then we have \( \mathbf{d} = (d_1, d_2, d_3, m, 0) \) and \( h = \eta \). Upon \( k \)-space averaging, denoted by \( \langle \cdots \rangle \), \( \tilde{H} \to m\Gamma^4 + \eta\Gamma^\mu \) and \( \mathbf{d}_\perp \cdot \Gamma^\perp \to 0 \) in Eq. D.12, yielding

\[
G_{00}^0(\omega) = a(\omega) + b_1(\omega)\Gamma^4 + b_2(\omega)\Gamma^\mu , \tag{D.13}
\]

with

\[
a(\omega) = \left\langle \frac{\tilde{\omega}(\tilde{\omega}^2 - d^2 - \eta^2)}{(\tilde{\omega}^2 - d^2 - \eta^2)^2 - 4\eta^2 d^2_\perp} \right\rangle = \frac{1}{2} \left\langle \frac{\tilde{\omega}}{\tilde{\omega}^2 - \tilde{E}_+^2} + \frac{\tilde{\omega}}{\tilde{\omega}^2 - \tilde{E}_-^2} \right\rangle,
\]

\[
b_1(\omega) = \left\langle \frac{m(\tilde{\omega}^2 - d^2 - \eta^2)}{(\tilde{\omega}^2 - d^2 - \eta^2)^2 - 4\eta^2 d^2_\perp} \right\rangle = \frac{1}{2} \left\langle \frac{m}{\tilde{\omega}^2 - \tilde{E}_+^2} + \frac{m}{\tilde{\omega}^2 - \tilde{E}_-^2} \right\rangle,
\]

\[
b_2(\omega) = \eta \left\langle \frac{\tilde{\omega}^2 - d^2 - \eta^2 + 2d^2_\perp}{(\tilde{\omega}^2 - d^2 - \eta^2)^2 - 4\eta^2 d^2_\perp} \right\rangle = \frac{\eta}{2} \left\langle \frac{1}{\tilde{\omega}^2 - \tilde{E}_+^2} + \frac{1}{\tilde{\omega}^2 - \tilde{E}_-^2} + \frac{4d^2_\perp}{(\tilde{\omega}^2 - \tilde{E}_+^2)(\tilde{\omega}^2 - \tilde{E}_-^2)} \right\rangle \tag{D.14}
\]

where \( \tilde{E}_+^2 = (E - \xi)^2 = d^2_\parallel + (d_\perp \pm \eta)^2 \), and

\[
\tilde{\omega} = \omega - \xi(k), \quad d^2 = \sum_{i=1}^3 d_i(k)^2 + m(k)^2, \quad d^2_\perp = \sum_{i=1}^3 d_i(k)^2(1 - \delta_{\mu,i}) . \tag{D.15}
\]

(2) if \( \Gamma = \Gamma^\mu, \mu \neq 4 \), (see Sec. 7.5.2), then \( \mathbf{d} = \eta \mathbf{e}_\mu + (d_1, d_2, d_3, m, 0) \) and \( h = 0 \). Upon \( k \)-averaging, Eq. D.12 is effectively \( M = \tilde{\omega}(\tilde{\omega}^2 - d^2) + (\tilde{\omega}^2 - d^2)(m\Gamma^4 + \eta\Gamma^\mu) \), thus

\[
G_{00}^0(\omega) = a(\omega) + b_1(\omega)\Gamma^4 + b_2(\omega)\Gamma^\mu \tag{D.16}
\]

where

\[
a(\omega) = \left\langle \frac{\tilde{\omega}}{\tilde{\omega}^2 - d^2} \right\rangle, \quad b_1(\omega) = \left\langle \frac{m}{\tilde{\omega}^2 - d^2} \right\rangle, \quad b_2(\omega) = \eta \left\langle \frac{1}{\tilde{\omega}^2 - d^2} \right\rangle \tag{D.17}
\]

with

\[
\tilde{\omega} = \omega - \xi(k), \quad d^2 = \sum_{i=1}^3 d_i(k)^2 + m^2 + \eta^2 . \tag{D.18}
\]
(3) If $\tilde{\Gamma} = \Gamma^{\mu\nu}$, $\mu \neq \nu \neq 4$ (see Sec. 7.5.3), then $d = (d_1, d_2, d_3, m, 0)$ and $h = \eta$. Upon $k$-space average, $\tilde{H} \rightarrow m\Gamma^4 + \eta\Gamma^{\mu\nu}$, $d_\perp \cdot \Gamma^\perp = m\Gamma^4$, $d_\mu \Gamma^\mu$ and $d_\nu \Gamma^{\nu} \rightarrow 0$, thus

$$G^0_{\eta\eta}(\omega) = a(\omega) + b_1(\omega)\Gamma^4 + b_2(\omega)\Gamma^{\mu\nu} + b_3(\omega)\Gamma^4\Gamma^{\mu\nu}$$  \hspace{1cm} (D.19)$$

where

$$a(\omega) = \left\langle \frac{\tilde{\omega}(\tilde{\omega}^2 - d^2 - \eta^2)}{(\tilde{\omega}^2 - d^2 - \eta^2)^2 - 4\eta^2d_\perp^2} \right\rangle = \frac{1}{2} \left\langle \frac{\tilde{\omega}}{\tilde{\omega}^2 - \tilde{E}_+^2} + \frac{\tilde{\omega}}{\tilde{\omega}^2 - \tilde{E}_-^2} \right\rangle$$

$$b_1(\omega) = \left\langle \frac{m(\tilde{\omega}^2 - d^2 + \eta^2)}{(\tilde{\omega}^2 - d^2 - \eta^2)^2 - 4\eta^2d_\perp^2} \right\rangle$$

$$b_2(\omega) = \eta \left\langle \frac{\tilde{\omega}^2 - d^2 - \eta^2 + 2d_\perp^2}{(\tilde{\omega}^2 - d^2 - \eta^2)^2 - 4\eta^2d_\perp^2} \right\rangle$$

$$b_3(\omega) = 2\eta \left\langle \frac{\tilde{\omega}m}{(\tilde{\omega}^2 - d^2 - \eta^2)^2 - 4\eta^2d_\perp^2} \right\rangle = 2\eta \left\langle \frac{\tilde{\omega}m}{(\tilde{\omega}^2 - \tilde{E}_+^2)(\tilde{\omega}^2 - \tilde{E}_-^2)} \right\rangle$$  \hspace{1cm} (D.20)$$

with $\tilde{E}_\pm^2 = (E - \xi)^2 = d_{\parallel}^2 + (d_\perp \pm \eta)^2$, and

$$\tilde{\omega} = \omega - \xi(k), d^2 = \sum_{i=1}^{3} d_i(k)^2 + m(k)^2,$$  \hspace{1cm} (D.21)$$

and

$$d_\perp^2 = \sum_{i=1}^{3} d_i(k)^2(1 - \delta_{\mu,i} - \delta_{\nu,i}) + m(k)^2(1 - \delta_{\mu,4} - \delta_{\nu,4}).$$  \hspace{1cm} (D.22)$$

### D.2 Green’s function of the BHB theory with prototypical $\tilde{\Gamma}$

In this section we use the following convention for $\Gamma$ matrices,

$$\Gamma^1 = I_x \otimes \sigma_x, \quad \Gamma^2 = I_x \otimes \sigma_y,$$  \hspace{1cm} (D.23)$$

$$\Gamma^3 = \tau_x \otimes \sigma_z, \quad \Gamma^4 = \tau_y \otimes \sigma_z, \quad \Gamma^5 = \tau_z \otimes \sigma_z,$$  \hspace{1cm} (D.24)$$
which is related to the one used in the text by a unitary transformation.

D.2.1 WS with point nodes ($\tilde{\Gamma} = \Gamma^{21}$)

The unperturbed Hamiltonian is

$$H(k) = \sum_{i=1}^{3} d_i(k_i) \Gamma^i + m \Gamma^4 + \eta \Gamma^{21}$$  \hspace{1cm} (D.25)

Taking $d_i(k_i) = k_i$ will give the linearized BHB Hamiltonian. Explicitly,

$$H(k) = I_\tau \otimes h_\sigma + h_\tau \otimes \sigma_z$$ \hspace{1cm} (D.26)

$$h_\sigma(k_x, k_y) = d_1(k_x) \sigma_x + d_2(k_y) \sigma_y \quad , \quad h_\tau(k_z) = d_3(k_z) \tau_x + m \tau_y + \eta I_\tau .$$ \hspace{1cm} (D.27)

Diagonalizing $h_\tau$ brings $H$ into block-diagonal form,

$$U^\dagger h_\tau U = \eta I_\tau + \sqrt{d_3^2 + m^2} \tau_z ,$$ \hspace{1cm} (D.28)

$$\tilde{H} \equiv U^\dagger H U = \begin{pmatrix} H_+ & \cr \cr \cr \end{pmatrix}$$ \hspace{1cm} (D.29)

where $U(k_z)$ and $U(k_z)$ are unitary matrices acting on the $\tau$ space and the $\tau \otimes \sigma$ space, respectively,

$$U(k_z) = \exp \left( -\frac{i\phi(k_z)}{2} \tau_z \right) \exp \left( -\frac{i\pi}{4} \tau_y \right) ,$$ \hspace{1cm} (D.30)

$$U(k_z) = U(k_z) \otimes I_\sigma ,$$ \hspace{1cm} (D.31)

$$\phi(k_z) = \tan^{-1} \frac{m}{d_3(k_z)} ,$$ \hspace{1cm} (D.32)

and the diagonal blocks of $\tilde{H}$ are labeled by $\tau = \pm 1$ with

$$H_\tau = B^\tau \cdot \sigma \quad , \quad B^\tau = (d_1, d_2, \eta + \tau \sqrt{d_3^2 + m^2}) \quad , \quad \tau = \pm 1 .$$ \hspace{1cm} (D.33)

The eigenvalues of $H$ are thus $\pm E_\tau$,

$$E_\tau = |B^\tau| = \sqrt{d^2 + m^2 + \eta^2 + 2\tau \eta \sqrt{d_3^2 + m^2}} .$$ \hspace{1cm} (D.34)

Note that for nonzero $\eta$ and $m$, bands with different $\tau$ indices can never cross.
Weyl nodes only exist in the $\tau = -1$ subspace in which the two bands touch at
\[ d = (0, 0, \pm \Delta) \quad , \quad \Delta = \sqrt{\eta^2 - m^2} . \] (D.35)

In the vicinity of $d_3 = \pm \Delta$, one writes
\[ d_3 = c(\Delta + q) \quad , \quad c = \pm 1 , \] (D.36)
then for $q \ll \eta$,
\[ H_-(k) = d_1 \sigma_x + d_2 \sigma_y - \Delta \frac{q}{\eta} \sigma_z + O\left(\frac{q^2}{\eta^2}\right) , \] (D.37)
and its spectrum is
\[ \lambda = \pm \sqrt{d_1^2 + d_2^2 + q^2 \Delta^2 \eta^2} . \] (D.38)

Note that $\Delta/\eta$ is related to the $\phi$ angle of the Weyl nodes via
\[ \cos \phi_c = \frac{c \Delta}{\eta} \quad , \quad \phi_c \equiv \phi|_{d_3 = c\Delta} . \] (D.39)

The local Green’s function is
\[ G^0_{00}(\omega) = \left\langle \mathcal{U}(k_z) \left( \begin{array}{c} G_+(\omega, k) \\ G_-(\omega, k) \end{array} \right) \mathcal{U}^\dagger(k_z) \right\rangle \] (D.40)
where $\langle \cdots \rangle$ denotes $k$-space average, and
\[ G_\tau(\omega, k) = \frac{\omega \mathbb{1}_\sigma + B_\tau^\ast(k_z) \sigma_z}{\omega^2 - |B_\tau(k)|^2} \quad , \quad \tau = \pm 1 \] (D.41)

$G_\tau$ is obtained from the Green’s function of $H_\tau(k)$ by dropping terms odd in $k_x$ and $k_y$ which would have averaged to zero.

Now we turn to the linearized theory $d_i(k_i) = k_i$. The aim is to isolate the contribution to the impurity effect near the nodal energy $\omega \sim 0$, from states near the Weyl nodes. The following approximations will be made:

1. We reduce the full $k$-space to two spheres of radius $Q$ around the two Weyl nodes labeled by their chirality $c = \pm 1$: $k = (k_x, k_y, c(\Delta + q))$ for $k_x, k_y, q \in [-Q, Q]$. In other words, the $k$-space average $\int d^3k \to \sum_{c=\pm 1} \int dk_x dk_y dq$. 
2. Within these spheres we will approximate \( U(k_z) \) by \( U(c \Delta) \), i.e., its value on the nodes, which is then moved out of \( \langle \cdots \rangle \) in Eq. D.40.

3. Furthermore, since the \( \tau = 1 \) eigenstates are gapped, \( G_-(\omega,k) \gg G_+(\omega,k) \) so in Eq. D.40 one can set \( G_+ = 0 \), i.e., project onto the \( \tau = -1 \) subspace.

Under these approximations the local Green’s function becomes

\[
G_{00}^0(\omega) = \left[ \sum_{c=\pm 1} U(c \Delta) \begin{pmatrix} 0 \\ 1 \end{pmatrix} \right] \otimes \langle G_-(\omega,k) \rangle \quad (D.42)
\]

\[
= G_0^0(\omega) (\mathbb{I}_\tau - \sin \phi_+ \tau_y) \otimes \mathbb{I}_\sigma , \quad (D.43)
\]

where \( \phi_+ \) is the \( \phi \) angle on the positive chirality node,

\[
\sin \phi_+ = \frac{m}{\eta} \quad (D.44)
\]

and

\[
G_0^0(\omega) \equiv \langle G_-(\omega,k) \rangle = \int \frac{dk_x dk_y dq}{(2\pi)^3} \frac{\omega - q \frac{\Delta}{\eta} \sigma_z}{\omega^2 - k_x^2 - k_y^2 - q^2 \frac{\Delta^2}{\eta^2}} . \quad (D.45)
\]

Note that \( \langle G_-(\omega,k) \rangle \) is proportional to \( \mathbb{I}_\sigma \) because the coefficient of \( \sigma_z \) is odd and integrates to zero. Introducing

\[
x = \frac{m}{\eta} \cos \theta , \quad u(x) = 1 - x^2 , \quad \kappa = \sqrt{u} \sqrt{k_x^2 + k_y^2 + q^2} , \quad (D.46)
\]

and using

\[
\int_0^K \frac{\kappa^2 d\kappa}{\omega^2 - \kappa^2} = -K + \frac{\omega}{2} \log \frac{\omega + K}{\omega - K} \quad (D.47)
\]

one has

\[
G_0^0(\omega) = \frac{\omega}{2\pi^2 m} \int_0^{\frac{m}{\pi}} \frac{dx}{u\sqrt{u}} \int_0^K \frac{\kappa^2 d\kappa}{\omega^2 - \kappa^2} \quad (D.48)
\]

\[
= \frac{Q \omega}{4\pi^2 \sin \phi_+} \log \frac{1 - \sin \phi_+}{1 + \sin \phi_+} + \frac{\omega^2}{4\pi^2 \sin \phi_+} \int_0^{\phi_+} d(\tan \phi) \log \frac{\omega + Q \cos \phi}{\omega - Q \cos \phi} , \quad (D.49)
\]

where we have used \( \sin \phi_+ = m/\eta \) and introduced \( \phi = \sin^{-1} x \). In the limit \( |\omega| \ll Q \), the second integral becomes \(-i\pi \tan \phi_+ \) (using \( \omega \to \omega + i0^+ \)).
D.2.2 WS with nodal line ($\tilde{\Gamma} = \Gamma^{35}$)

Consider the Hamiltonian

$$H(k) = \sum_{i=1}^{3} d_i(k_i) \Gamma^i + m \Gamma^4 + \eta \Gamma^{35} \quad \text{(D.50)}$$

$$= I_\tau \otimes (d_1 \sigma_x + d_2 \sigma_y) + \tau_y \otimes (\eta I_\sigma + m \sigma_z) + d_3 \tau_x \otimes \sigma_z \quad \text{(D.51)}$$

To block diagonalize, we first rotate $(\tau_y, \tau_z) \rightarrow (\tau_z, -\tau_y)$, and then send, simultaneously, $\tau_\pm \otimes I_\sigma \rightarrow \tau_\pm \otimes \sigma_z$ and $I_\tau \otimes \sigma_\pm \rightarrow \tau_\pm \otimes \sigma_\pm$, which is the unitary transformation $U = \text{diag}(1, 1, 1, -1)$. This is equivalent to taking the following $\Gamma$ matrix convention from the outset (switching the order of $\tau$ and $\sigma$ spaces in the direct product),

$$\Gamma^1 = \sigma_x \otimes \tau_z \quad , \quad \Gamma^2 = \sigma_y \otimes \tau_z \quad , \quad \Gamma^3 = I_\sigma \otimes \tau_x \quad , \quad \Gamma^4 = \sigma_z \otimes \tau_z \quad , \quad \Gamma^5 = -I_\sigma \otimes \tau_y \quad \text{(D.52)}$$

After this basis change, one has

$$H(k) = h_\sigma \otimes \tau_z + d_3 I_\sigma \otimes \tau_x \quad , \quad h_\sigma = d_1 \sigma_x + d_2 \sigma_y + m \sigma_z + \eta I_\sigma \quad \text{(D.54)}$$

Diagonalizing $h_\sigma$ then brings $H(k)$ into block-diagonal form,

$$U^\dagger h_\sigma U = \eta I_\sigma + \sqrt{d_1^2 + d_2^2 + m^2} \sigma_z \quad \text{(D.55)}$$

$$\tilde{H} \equiv U^\dagger H U = \begin{pmatrix} H_+ \\ H_- \end{pmatrix} \quad \text{(D.56)}$$

where $U(k_x, k_y)$ and $U(k_x, k_y)$ are unitary matrices acting on the $\sigma$ space and the $\sigma \otimes \tau$ space, respectively,

$$U(k_x, k_y) = \exp \left( -\frac{i}{2} \phi(k_x, k_y) \sigma_z \right) \exp \left( -\frac{i}{2} \theta(k_x, k_y) \sigma_y \right) \quad , \quad \text{(D.57)}$$

$$U(k_x, k_y) = U(k_x, k_y) \otimes I_\tau \quad , \quad \text{(D.58)}$$

and $\theta$, $\phi$ are the polar and azimuthal angles of the vector $(d_1, d_2, m)$. The diagonal blocks of $\tilde{H}$ are labeled by $s = \pm 1$ with

$$H_s = d_3 \tau_x + \left( \eta + s \sqrt{d_1^2 + d_2^2 + m^2} \right) \tau_z \quad , \quad s = \pm 1 \quad \text{(D.59)}$$
The eigenvalues of $H$ are thus $\pm E_s$,

$$E_s = \sqrt{d_3^2 + \left( \eta + s \sqrt{d_1^2 + d_2^2 + m^2} \right)^2}.$$  \hspace{1cm} (D.60)

Note that for nonzero $\eta$ and $m$, bands with different $s$ indices can never cross.

Weyl nodes only exist in the $s = -1$ subspace in which the two bands touch at

$$d = (\Delta \cos \phi , \Delta \sin \phi , 0) \quad , \quad \Delta = \sqrt{\eta^2 - m^2}.$$  \hspace{1cm} (D.61)

In the vicinity of the line node, one writes

$$d_1 = (\Delta + q) \cos \phi \quad , \quad d_2 = (\Delta + q) \sin \phi ,$$  \hspace{1cm} (D.62)

then for $q \ll \eta$,

$$H_-(k) = d_3 \tau_x - \Delta \frac{q}{\eta} \tau_z + O \left( \frac{q^2}{\eta^2} \right) ,$$  \hspace{1cm} (D.63)

and its spectrum is

$$\lambda = \pm \sqrt{d_3^2 + q^2 \Delta^2 \frac{\eta^2}{\eta^2}}.$$ \hspace{1cm} (D.64)

Note that $\Delta/\eta$ is related to the polar angle $\theta$ of the vector $(d_1, d_2, m)$ on the nodal line,

$$\sin \theta_N = \frac{\Delta}{\eta}$$ \hspace{1cm} (D.65)

The local Green’s function is

$$G_{00}^0(\omega) = \left\langle U(k_x, k_y) \begin{pmatrix} G_+(\omega, k) \\ G_-(\omega, k) \end{pmatrix} U^\dagger(k_x, k_y) \right\rangle$$  \hspace{1cm} (D.66)

where $\langle \cdots \rangle$ denotes $k$-space average, and

$$G_s(\omega, k) = \frac{\omega \tau_x + \left( \eta + s \sqrt{d_1^2 + d_2^2 + m^2} \right) \tau_z}{\omega^2 - d_3^2 - \left( \eta + s \sqrt{d_1^2 + d_2^2 + m^2} \right)^2} , \quad s = \pm 1 .$$  \hspace{1cm} (D.67)
\( G_s \) is obtained from the Green’s function of \( H_s(k) \) by dropping terms odd in \( k_z \), i.e. the one proportional to \( \tau_x \) in the numerator, which would have averaged to zero.

Now we turn to the linearized theory \( d_i = k_i \) and investigate the contribution of states near the line Weyl node to the impurity effect near zero energy \( \omega \sim 0 \). We employ the following approximations,

1. The full \( k \)-space is reduced to a tube of radius \( Q \) around the line node, \( k = ((\Delta + q) \cos \phi, (\Delta + q) \sin \phi, k_z) \) with \( \sqrt{q^2 + k_z^2} \in [0, Q] \). The \( k \)-space average \( \int d^3k \to (\int d\phi/2\pi)(\int 2\pi dq dk_z) \).

2. Within the tube we will approximate \( U(k_x, k_y) \) by its value on the nodal line, \( U(\Delta \cos \phi, \Delta \sin \phi) \). It is then taken out of the average over the tube’s cross-section.

3. Since the \( s = 1 \) states are gapped, \( G_- (\omega, k) \gg G_+ (\omega, k) \) for \( \omega \ll \eta \), so in Eq. D.66 one can set \( G_+ = 0 \), i.e., project onto the \( s = -1 \) subspace.

Under these approximations, the local Green’s function becomes

\[
G^0_{00} = \left\langle U(\Delta, \phi) \begin{pmatrix} 0 \\ 1 \end{pmatrix} U^\dagger (\Delta, \phi) \right\rangle \otimes \left\langle G_-(\omega, k) \right\rangle_{q, k_z} \quad (D.68)
\]

\[
= \frac{\Delta}{2} G^0_- (\omega) (\mathbb{I}_\sigma - \cos \theta_N \sigma_z) \otimes \mathbb{I}_\tau , \quad (D.69)
\]

where \( \theta_N \) is the aforementioned polar angle of the nodal line in the \((k_x, k_y, m)\) space,

\[
\cos \theta_N = \frac{m}{\eta} \quad (D.70)
\]

and

\[
G^0_-(\omega) \equiv \left\langle G_-(\omega, k) \right\rangle_{q, k_z} = \int_0^Q dq \, dk_z \frac{\omega - q \frac{\Delta}{\eta} \sigma_z}{(2\pi)^2 \omega^2 - k_z^2 - q^2 \frac{\Delta^2}{\eta^2}} . \quad (D.71)
\]

Note that \( G^0_-(\omega) \) is a number because the coefficient of \( \sigma_z \) is odd in \( q \) and integrates to zero. Denoting

\[
\tan \chi = \frac{q}{k_z} , \quad u(\chi) = 1 - \frac{m^2}{\eta^2} \sin^2 \chi , \quad \kappa = \sqrt{u} \sqrt{k_z^2 + q^2} , \quad (D.72)
\]
one has

\[ G_0^-(\omega) = \frac{\omega}{4\pi^2} \int_0^{2\pi} \frac{d\chi}{u} \int_0^\infty \frac{\kappa d\kappa}{\omega^2 - \kappa^2} \]  

\[ = -\frac{\omega}{2\pi^2} \int_0^{\pi/2} \frac{d\chi}{u} \log \left[ 1 - u \frac{Q^2}{\omega^2} \right] . \]  

(D.73)  

(D.74)

In the limit \( \omega \ll Q \),

\[ \log(1 - uQ^2/\omega^2) \simeq \log \left[ -uQ^2/\left(\omega + i0^+\right)^2 \right] = \log(uQ^2/\omega^2) + i\pi \text{sgn}(\omega) , \]  

(D.75)

thus

\[ G_0^-(\omega) = \left( \frac{Q}{\pi^2} \cdot \frac{\omega}{Q} \log \frac{|\omega|}{2\pi} - i \frac{|\omega|}{2\pi} \right) \int_0^{\pi/2} \frac{d\chi}{u(\chi)} - \frac{\omega}{2\pi^2} \int_0^{\pi/2} \frac{d\chi}{u(\chi)} \log u(\chi) \]  

\[ = R(\theta_N) \omega - i \frac{|\omega|}{4 \sin \theta_N} , \]  

(D.76)  

(D.77)

with

\[ R(\theta_N) = - \int_0^{\pi/2} \frac{d\chi}{2\pi^2} \frac{\log(1 - \cos^2 \theta_N \sin^2 \chi)}{1 - \cos^2 \theta_N \sin^2 \chi} . \]  

(D.78)
Bibliography


