TOPOLOGICAL MATERIALS: PHASE TRANSITIONS AND MAGNETOELECTRIC RESPONSE

BY JIANPENG LIU

A dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Physics and Astronomy

Written under the direction of

David Vanderbilt

and approved by

_____

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2015

ABSTRACT OF THE DISSERTATION

Topological materials: phase transitions and magnetoelectric response

by Jianpeng Liu

Dissertation Director: David Vanderbilt

In this thesis, we study the properties of topological materials using theoretical techniques such as first-principles calculations and tight-binding models. In the first part of the thesis, we deal with the phase transitions in 3D topological insulators. We first study the topological phase transitions in In- and Sb-doped Bi_2Se_3, where distinct behaviors are found. In the In-doped case, we find that the In $5s$ orbitals destroy the topological phase at low impurity compositions, and the phase transition is better described by a local percolation scenario. On the other hand, the Sb-doped Bi_2Se_3 is well described by a "linear-gap-closure" picture, where the phase transition is dominated by the gradual decrease of the effective spin-orbital coupling. We also discuss the Weyl semimetals emerging from noncentrosymmetric topological insulators. We first clarify the previous theory, and prove that an intermediate Weyl semimetal must show up through the phase transition from a 3D topological to normal insulator. Then we propose $LaBi_{1-x}Sb_xTe_3$, $LuBi_{1-x}Sb_xTe_3$ and pressurized BiTeI as possible candidates of Weyl semimetals.

The second part of the thesis is focused on method development. We first propose

a quantitative definition for the band inversions driven by spin-orbit coupling in insulators, known as the "spin-orbit spillage". The spin-orbit spillage has been applied to various topological systems, which turns out to be a useful tool for the identification of topological characters in band theory. In the last chapter of the thesis, we develop a new method for calculating the Chern-Simons orbital magnetoelectric coupling in 3D insulators. The contributions from the gauge discontinuity and the "vortex loops" are taken into account in our method. The former is expressed as a 2D integral over a $\mathbf{k}$ plane across which the gauge of the occupied Bloch functions become discontinuous, while the latter is expressed as the Berry phases around 1D "vortex loops" lying in the gauge-discontinuity plane. Our method is successfully applied to the Fu-Kane-Mele model with the breaking of time-reversal symmetry.

Acknowledgements

The five years at Rutgers would definitely be one of the best times in my life. I have received a solid physics education here, and have evolved from a student to a physics researcher. I have met so many nice people in these five years, and I would never expect to finish this thesis without their help.

First I would like to thank those professors with whom I have taken my graduate courses, including Natan Andrei, Piers Coleman, Kristjan Haule, Lev Ioffe, Gabriel Kotliar, Joel Lebowitz, Stephen Schnetzer, Joel Shapiro and David Shih. I have taken three courses with Joel Shapiro, from which I have got the training as a theoretical physicist; and I have taken two courses with Natan Andrei, who introduced modern solid state physics to me. I am also impressed by the many-body-physics course with Piers Coleman, not only by the wonderful lectures but also by his passion to physics. I enjoyed the discussions and collaborations with Seongshik Oh and Weida Wu, from which I have learned some basic knowledge about experimental condensed matter physics. I am especially grateful to Karin Rabe, who gave me tons of advices for presentations, and served as one of my references; to Premala Chandra, who helped me to make a tough decision in my academic career; and to Ivo Souza, who was also one of my references, and helped me a lot in my research.

I also learned a lot from the discussions with postdoctoral fellows, graduate students and visitors at Rutgers, including Vitya Alexandrov, Sergey Artyukhin, Joseph Bennett, Turan Birol, Matthew Brahlek, Sinisa Coh, Kevin Garrity, Yanan Geng, Qiang Han, Jiawang Hong, Huaqing Huang, Bryan Leung, Wenshuo Liu, Se Young Park, Sebastian Reyes-Lillo, Quantong Shen, Alexey Soluyanov, Maryam Taherinejad, Wenbo

Wang, Wenhu Xu, Meng Ye, Chuck-Hou Yee, Tahir Yusufaly, Hongbin Zhang, Jialan Zhang, Wenhan Zhang, Yuanjun Zhou, Qibin Zhou, and Shucheng Zhu. I would like to particularly thank Kevin Garrity and Hongbin Zhang, who helped me a lot in the early stage of my research. Of course, this list is far from being complete.

All the projects I have done at Rutgers were under the supervision of my advisor David Vanderbilt. David led me to the field of topological phases of matter, and suggested quite a few exciting projects for me, from which I learned almost all the advanced physics I know now. As David's student, I am always encouraged to think about my own ideas, and to discuss and collaborate with others. All in all, I would like to express my sincere gratitude to David for being such an amazing advisor.

I also feel lucky that I have made a lot new friends in these five years. I would like to thank all the people who played basketball with me on weekends, without whom the life would be much less fun.

In the end, I extend my gratitude to my family, my wife Huijie Guan, my mother Guiqin Ji, my father Yangming Liu, my aunt Guirong Ji and my grandmother Xiue Wen, for their consistent love and support.

Dedication

To my family

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Condensed matter physics deals with systems consisted of a macroscopically large number of electrons and nuclei. The most generic Hamiltonian of a condensed matter system includes the kinetic energy of electrons and nuclei, the electron-electron (e-e) and electron-nuclei (e-n) Coulomb interactions, as well as the spin-orbit coupling as a relativistic effect in systems with heavy elements. Approximations have to be made in order to deal with systems with such complexity. For example, to deal with the electronic behavior, one can first isolate the nuclear dynamics from the electrons by freezing the nuclear degrees of freedom and consider a static periodic lattice structure, then treat the e-n Coulomb interaction as an external periodic potential. Such an approximation is usually valid because the characteristic time scale of nuclear dynamics is much slower than that of electrons.

If one further makes a drastic approximation that the e-e Coulomb interaction can also be treated on a mean-field level, i.e., the complicated interacting many-electron system is approximated as single electrons moving in an effective static field generated self-consistently by other electrons, then the problem is reduced to a simpler problem of solving for many non-interacting electrons obeying Pauli exclusion principle and moving in an effective static field with lattice periodicity. The eigenstates of such a mean-field effective Hamiltonian are known as Bloch states $\psi_{n\mathbf{k}}(\mathbf{r})$, where $\mathbf{k}$ is the wave vector which is periodic in the Brillouin zone, and n is the band index. The corresponding eigenenergies $\epsilon_{n\mathbf{k}}$ gives us the bandstructure. A Bloch eigenfunction is of the form $\psi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{n\mathbf{k}}(\mathbf{r})$, where $u_{n\mathbf{k}}(\mathbf{r})$ has the periodicity of the primitive unit cell. Such

a theory is known as the band theory of solids.

Band theory lays the foundation for modern solid state physics. For example, band theory has successfully explained the difference between a metal and an insulator: an insulator is characterized by an energy gap between the occupied and unoccupied Bloch states, while such a gap is absent in a metal. Moreover, the bandstructures and Bloch functions of a given material typically act as the cornerstone based on which much interesting physics in solids, such as electric transport phenomena, effects of vacancies and impurities, and optical properties etc., are developed [1].

In particular, with the development of first-principles techniques calculating electronic structure in solids [2], such as the density functional theory [3, 4], it becomes feasible to calculate properties of realistic materials in an accurate and "parameter-free" way. Based on first-principles band theory, modifications can be made to study many interesting effects in realistic materials. To name a few, using the iterative Green's function technique, one can study the surface and interface physics of crystals based on the Bloch states and bandstructures of the bulk materials [5, 6, 7]; an onsite Hubbard-like Coulomb interaction treated on the Hartree-Fock level can be added on top of the Kohn-Sham [4] type Hamiltonian to study the correlated physics in transition metal oxides [8]; using the "maximally localized Wannier functions" [9] obtained from first-principles Bloch functions, various response functions in solids such as the dc electric conductivity tensor and magneto-optical absorption can be evaluated with great accuracy and numeric efficiency [10]. A plethora of research directions thrive by virtue of the band theory and the development of first-principles methods in calculating electronic structures of solids.

Besides the interesting physics mentioned above, the effects of the geometric phase in crystalline solids have attracted a significant amount of interest since the discovery of Berry's geometric phase in quantum mechanics [11, 12]. In crystalline solids, the wavevector $\mathbf{k}$ is periodic in the Brillouin zone (BZ). A geometric phase thus can be

defined for a Bloch function which varies adiabatically on a $\mathbf{k}$ path in the toric BZ. Such a Berry phase accumulated by varying Bloch functions in the BZ was first discussed by Zak, and sometimes known as Zak's phase [13].

Various interesting phenomena in solids are related to the Berry phase. For example, electric polarizations in band insulators stems from the Berry phase of Bloch functions in the BZ [14, 15]; the intrinsic contribution to the anomalous Hall effect in (quasi)2D systems [16] originates from the Berry phase of the Bloch function around the Fermi loop [17, 18]; and the Chern-Simons orbital magnetoelectric coupling turns out to be contributed by a more complicated geometric term, i.e., the 3D BZ integral of the Chern-Simons 3-form [19, 20].

For band insulators in even dimensions, an integer topological index called the Chern number can be defined in the corresponding even-dimensional BZ. For a "zero-dimensional" molecule or cluster, the zeroth Chern number is simply the number of occupied particles. For a 2D insulators, the first Chern number is defined as the integration of Berry curvature over the 2D BZ torus. There is also the second Chern number in four dimensions, but it is not quite relevant to physical observables. We will use the term "Chern number" to refer to the first Chern number defined in the 2D BZ throughout this thesis. A non-zero Chern number is responsible for a quantized Hall conductance in insulating 2D systems. It explains the integer quantum Hall (IQH) effect [21] from a topological perspective [22, 23] and has motivated research on the quantum anomalous Hall effect in (quasi) 2D material systems [24, 25].

It can be shown that the Chern number must vanish in the presence of time-reversal symmetry. However, in 2005, it was proposed by Kane and Mele that time-reversal ($\mathcal{T}$) symmetry would impose an additional $\mathbb{Z}_2$ topological classification to 2D band insulators [26]. An odd $\mathbb{Z}_2$ index corresponds to the quantum spin Hall (QSH) phase, which is manifested by an odd number of pairs of spin-polarized gapless [27, 26, 28] edge states which can never exist in isolated 1D systems [29]. An even $\mathbb{Z}_2$ index corresponds to the

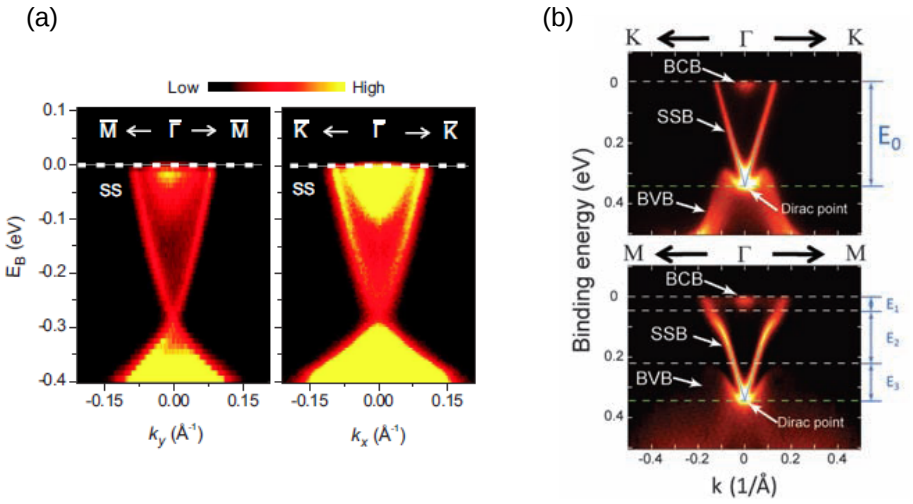

Figure 1.1: ARPES results for (a) Bi_2Se_3, and (b) Bi_2Te_3. Figure (a) is taken from Nature Physics, vol. 5, p. 398 (2009), and Fig. (b) is from Science, vol. 325, p. 178 (2009).

conventional 2D insulators without topologically protected edge states.

The $\mathbb{Z}_2$ classification of 2D insulators was soon generalized 3D $\mathcal{T}-$invariant systems in 2007 [30, 31]. In 3D systems, it turns out that there are four $\mathbb{Z}_2$ indices, where one of them is the known as the strong $\mathbb{Z}_2$ index and the rest of the three are weak indices [30, 31, 32]. A system with an odd strong $\mathbb{Z}_2$ index is now known as a "topological insulator", which has an odd number of gapless Dirac cones on the surface and possesses interesting transport properties and magnetoelectric response [33, 34].

Topological insulators (TIs) have been theoretically proposed and experimentally realized in various material systems [35]. The most famous one may be the Bi_2Se_3-class materials [36, 37, 38, 39]. The Bi_2Se_3-class materials, including Bi_2Se_3, Bi_2Te_3 and Sb_2Te_3, were first theoretically proposed as candidates of TIs in 2009 [36]. These theoretical predictions were later on experimentally verified by angle resolved photoelectron spectroscopy (ARPES) [37, 38, 39]. The ARPES results for Bi_2Se_3 and Bi_2Te_3 are shown in Fig. 1 (a) and (b) respectively. It is clearly seen that there is a single Dirac cone traversing the bulk energy gap centered at one of the TRIM (Γ) in both materials.

An interesting topic is the topological phase transitions in this class of topological insulators driven by impurity doping. One of the main topics of this paper is to study the topological phase transitions in In- and Sb-doped Bi_2Se_3 [40, 41, 42]. As will be discussed in detail in Chapter 3, distinct behaviors have been observed in the phase-transition process of these two solid-solution systems [42].

A topological index is usually well defined for systems with a non-vanishing bulk energy gap. However, recently the concept of topologically protected phases have been generalized to gapless systems with band-touching points (BTPs) in the BZ. Systems with such topologically protected BTPs are known as Weyl semimetals (WSMs) [43, 44, 45], which are topologically robust only in three dimensions. It turns out the emergence of the WSMs requires the breaking of either $\mathcal{T}$ symmetry or inversion ($\mathcal{P}$) symmetry. In particular, in noncentrosymmetric TIs, WSMs show up as a critical phase in the process of topological phase transitions connecting the $\mathbb{Z}_2$-odd and $\mathbb{Z}_2$-even insulating phases [46, 47, 48]. In Chapter 4, we will discuss the emergence of WSMs as such a critical phase. We also propose a couple of material candidates of WSMs based on noncentrosymmetric TIs [48].

A crucial ingredient in topological materials is spin-orbit coupling (SOC). In systems with weak electron-electron interactions [1], SOC is indispensable in obtaining topologically nontrivial phases. The general scenario is that SOC drives a band inversion somewhere in the BZ, and such a band inversion may lead to a nontrivial band topology. In Chapter 5, we define a quantity called "spin-orbit spillage" [51]. It allows for a quantitative evaluation of SOC-driven band inversions in the BZ, which may be helpful for search for new topological materials.

An important signature of 3D TIs is the quantized magnetoelectric (ME) response [19, 20]. To be specific, the Chern-Simons orbital ME response of 3D TIs is formally

[1] It is possible to have "topological Mott insulators" in strongly correlated systems [49]. In these systems, some kind of effective spin-orbit coupling is generated dynamically by electron-electron interactions [50], leading to nontrivial band topology. We will not discuss this topic in the present thesis.

quantized as $e^2/2h$ [2], which can be considered as a benchmark of TIs. The Chern-Simons orbital ME coupling is not only present in TIs, it exists in all insulating solids with the breaking of both $\mathcal{T}$ and $\mathcal{P}$ symmetry. For systems in the absence of $\mathcal{T}$ and $\mathcal{P}$ symmetries, it is no longer quantized and can take arbitrary values depending on specific material properties. However, the existing method for computing the Chern-Simons term suffers from the difficulty in numeric convergence with respect to **k**-mesh sampling. In Chapter 6, we propose a new method to calculate the orbital Chern-Simons ME response. We derive new formulas for the contributions from gauge discontinuity of Bloch functions in the BZ. It turns out that our method is numerically efficient and may provide a new interpretation to the origin of the quantized ME response in TIs.

This paper is organized as follows. In Chapter 2, we will review some basic concepts. We will start by introducing the concept of the Berry phase, then show a few examples of geometric-phase effects in solids. We proceed to discuss some basic topologically-protected phases in condensed matter physics. After Chapter 2, the thesis is divided into two parts. The first part includes topics related to Bi_2Se_3-class TIs, including the topological phase transitions in In- and Sb-doped Bi_2Se_3 (Chapter 3) and WSMs emerging from non-centrosymmetric TIs (Chapter 4). The second part is focused on method development. As mentioned above, in Chapter 5, the concept of spin-orbit spillage is proposed, and is applied to various topological systems. In Chapter 6, a new method for computing the Chern-Simons orbital ME coefficient is proposed. This method is successfully applied to the Fu-Kane-Mele model [30] with broken time-reversal symmetry.

[2]For a TI sample, the quantized bulk ME response is exactly cancelled by the contributions from the surface states such that the total ME response of the sample is still vanishing. We will discuss this topic in details in Sec. 2.2.4.

Chapter 2

Basic concepts

In this chapter, we introduce some basic concepts to understand topological phases of matter. First we introduce the concept of Berry's geometric phase, and discuss its applications in solids such as the electric polarization, the anomalous Hall effect and the Chern-Simons magnetoelectric coupling. Then we make a transition from geometry to topology by discussing a few examples of topologically protected phases. We start with the first example of topologically protected phases in condensed matter physics, the integer quantum Hall (IQH) effect, and continue to discuss the "quantized Hall effect without magnetic field", i.e., the quantum anomalous Hall (QAH) effect. Then we look at the situations with time-reversal symmetry. We will talk about the quantum spin Hall (QSH) phase in 2D insulators with time-reversal symmetry and its 3D generalization for topological insulators.

2.1 Berry phase

Let us consider a quantum system which is coupled to a classical environment parameterized by $\mathbf{R}(t) = \{R_1(t), R_2(t), ..., R_N(t)\}$. The parameters $\mathbf{R}$ are allowed to vary as a function of time t, but we assume that the time evolution is slow enough compared with any intrinsic dynamics of the quantum system, such that one can define a set of eigenenergies and orthonormal eigenstates of the Hamiltonian $H(\mathcal{R}(t))$, which are

instantaneously dependent on $\mathcal{R}(t)$

$$H(\mathcal{R}(t))|n(\mathcal{R}(t)\rangle = E_n(\mathbf{R}(t))|n(\mathbf{R}(t))\rangle \ ,$$
$$\langle m(\mathbf{R}(t))|n(\mathbf{R}(t))\rangle = \delta_{mn} \ . \tag{2.1}$$

We also assume that the phase factor, i.e., the "gauge" of $|n(\mathbf{R}(t))\rangle$ has been chosen in such a way that it remains smooth and single-valued in the parameter space $\mathbf{R}(t)$. Then we study the time evolution of an initial state at $t = 0$ ($|\psi(0)\rangle$) in response to the variation of the external parameters in a periodic cycle from $\mathbf{R}(0)$ to $\mathbf{R}(\tau) = \mathbf{R}(0)$.

Suppose that the system starts from one of its eigenstates $|\psi(0)\rangle = |n(\mathbf{R}(0))\rangle$, then in general the corresponding state at time t $(0 < t < \tau)$ $|\psi(\mathbf{R}(t))\rangle$ should be a superposition of all the instantaneous eigenstates

$$|\psi(\mathbf{R}(t))\rangle = \sum_m c_m(t)|m(\mathbf{R}(t))\rangle \ . \tag{2.2}$$

However, if the time evolution is slow enough, we can make the adiabatic approximation. The adiabatic approximation says that if the system starts in one of the eigenstates $|n(\mathbf{R}(0))\rangle$ at $t = 0$, then it remains in the corresponding instantaneous eigenstate $|n(\mathbf{R}(t))\rangle$ throughout the evolution process. In other words, in the adiabatic approximation,

$$|\psi(\mathbf{R}(t))\rangle = c_n(t)|n(\mathbf{R}(t))\rangle \ , \tag{2.3}$$

where $c_n(t)$ is a phase factor determined by the path of time evolution, with $c_n(0) = 1$. The adiabatic approximation is valid when the variation of $\mathbf{R}(t)$ is much slower than any of the intrinsic time scales of the system; or equivalently, the energy fluctuation induced by the time variation of $\mathbf{R}(t)$ is much smaller than typical energy gaps of the system.

Plugging Eq. (2.3) into the time-dependent Schrödinger equation, one would immediately obtain the following equation for $c_n(t)$,

$$\frac{dc_n(t)}{dt} = -c_n(t)\Big(iE_n(t) + \langle n(\mathbf{R}(t))|\frac{d}{dt}n(\mathbf{R}(t))\rangle\Big) \ . \tag{2.4}$$

Dividing both sides of Eq. (2.4) by $c_n(t)$, and integrating over time from 0 to t, one obtains,

$$c_n(t) = e^{-i\int_0^t E_n(\mathbf{R}(t'))dt'} e^{i\beta_n(t)} \tag{2.5}$$

where

$$\begin{aligned} \beta_n(t) &= \int_0^t i\,\langle n(\mathbf{R}(t'))|\frac{d}{dt'}n(\mathbf{R}(t'))\rangle dt' \\ &= \int_{\mathbf{R}(0)}^{\mathbf{R}(t)} i\,\langle n(\mathbf{R})|\frac{d}{d\mathbf{R}}n(\mathbf{R})\rangle d\mathbf{R}\ . \end{aligned} \tag{2.6}$$

One notices that besides the expected "dynamic" phase factor $e^{-i\int_0^t E_n(\mathbf{R}(t'))dt'}$, an extra phase factor $e^{i\beta_n(t)}$ resulted from the variation of the eigenbasis also plays a role. It was argued that such a phase factor is unimportant because it is gauge dependent, and it vanishes if a proper gauge choice is made. Then the only gauge-invariant phase factor is the dynamic phase. This is indeed true for a non-cyclic evolution. On the other hand, for cyclic adiabatic evolution (with period τ), the phase factor $e^{i\beta(\tau)}$ becomes gauge invariant (modulo 2π). Therefore, for a cyclic evolution,

$$\begin{aligned} \psi(\tau) &= e^{-i\int_0^\tau E_n(\mathbf{R}(t'))dt'} e^{i\beta_n(\tau)}|n(\mathbf{R}(\tau))\rangle \\ &= e^{-i\int_0^\tau E_n(\mathbf{R}(t'))dt'} e^{i\beta_n(\tau)}|n(\mathbf{R}(0))\rangle\ . \end{aligned} \tag{2.7}$$

The second line in the above equation follows as a result of the single-valued gauge choice of $|n(\mathbf{R})\rangle$. We see that for a cyclic adiabatic evolution, a system returns to a state which differs from the original state by an extra phase factor in addition to the conventional dynamic phase. Such a phase factor is known as the Berry phase. The Berry phase is important because (unlike the case of non-cyclic evolution) it is gauge invariant modulo 2π, implying that it may be related to physical observables.

Now let us illustrate the gauge invariance of the Berry phase. For a cyclic evolution, the system traverses a closed path $\mathcal{C}$ in the parameter space $\mathbf{R}$. From Eq. (2.6) we know the Berry phase can be expressed as

$$\beta_n = \oint_{\mathcal{C}} i\,\langle n(\mathbf{R})|\frac{d}{d\mathbf{R}}n(\mathbf{R})\rangle \cdot d\mathbf{R}\ . \tag{2.8}$$

The integrand on the right-hand side (RHS) of the above equation is known as the Berry connection $\mathbf{A}^n(\mathbf{R})$:

$$\mathbf{A}^n(\mathbf{R}) = i\langle n(\mathbf{R})|\frac{d}{d\mathbf{R}}n(\mathbf{R})\rangle \ . \tag{2.9}$$

Suppose a $U(1)$ gauge transformation is made to $|n(\mathbf{R})\rangle$ such that

$$|n(\mathbf{R})\rangle \to e^{i\lambda_n(\mathbf{R})}|n(\mathbf{R})\rangle \ , \tag{2.10}$$

where $e^{i\lambda_n(\mathbf{R})}$ is single-valued on $\mathbf{R}$. Then it is straightforward to show that

$$\begin{aligned} &\mathbf{A}^n(\mathbf{R}) \to \mathbf{A}^n(\mathbf{R}) - d\lambda_n(\mathbf{R})/d\mathbf{R} \\ &\beta_n \to \beta_n - \lambda_n(\mathbf{R}(\tau)) + \lambda_n(\mathbf{R}(0)). \end{aligned} \tag{2.11}$$

As $e^{i\lambda_n(\mathbf{R})}$ is single-valued, it immediately follows that $\lambda_n(\mathbf{R}(\tau)) - \lambda_n(\mathbf{R}(0)) = 2\pi n$ with n to be an arbitrary integer. Thus it is proved that the Berry phase is gauge invariant modulo 2π.

The loop integral shown in Eq. (2.8) can also be expressed as an area integral of a local gauge-invariant quantity known as the Berry curvature. By applying Stoke's theorem to the loop integral in Eq. (2.8), one obtains

$$\begin{aligned} \beta_n &= \oint_{\mathcal{C}} \mathbf{A}^n(\mathbf{R}) \cdot d\mathbf{R} \\ &= \oint_{\mathcal{C}} \sum_{i=1}^{N} A_i^n(\mathbf{R}) dR_i \\ &= \int_{\mathcal{S}} \sum_{i,j=1}^{N} \frac{1}{2}\Omega_{ij}^n dR_i \wedge dR_j \end{aligned} \tag{2.12}$$

where $\mathcal{S}$ is a subregion in the multi-dimensional parameter space $\mathbf{R}$ enclosed by loop $\mathcal{C}$, and "$\wedge$" denotes the wedge product $dR_i \wedge dR_j = -dR_j \wedge dR_i$. Ω_{ij}^n is the Berry curvature defined as

$$\begin{aligned} \Omega_{ij}^n &= \partial A_j^n/\partial R_i - \partial A_i^n/\partial R_j \\ &= i\,\langle \partial_{R_i} n|\partial_{R_j} n\rangle - i\langle \partial_{R_j} n|\partial_{R_i} n\rangle \ . \end{aligned} \tag{2.13}$$

In Eq. (2.12) and Eq. (2.13), the dependence on $\mathbf{R}$ is implicit (in later discussions, we will drop the explicit dependence of $\mathbf{R}$). As a simplest example, in 2D parameter space, $\mathbf{R} = (R_1, R_2)$,

$$\begin{aligned} \beta_n &= \int_{\mathcal{S}} \frac{1}{2}\Big(\Omega^n_{12} dR_1 \wedge dR_2 + \Omega^n_{21} dR_2 \wedge dR_1 \Big) \\ &= \int_{\mathcal{S}} \Omega^n_{12} dR_1 \wedge dR_2 \ , \end{aligned} \tag{2.14}$$

where we have used the antisymmetric property of the wedge product and the Berry curvature.

Unlike the Berry connection, it is straightforward to check that the Berry curvature is a gauge invariant quantity. Moreover, as a result of

$$\langle m|\partial_{R_i} n\rangle = \frac{\langle m|\partial H/\partial R_i|n\rangle}{E_n - E_m} \ , \ \text{ if } m \neq n \ , \tag{2.15}$$

the Berry curvature can be rewritten in a form that resembles a response function obtained from perturbation theory,

$$\Omega^n_{ij} = i \sum_{m\neq n} \frac{\langle n|\partial H/\partial R_i|m\rangle\langle m|\partial H/\partial R_j|n\rangle - \langle n|\partial H/\partial R_j|m\rangle\langle m|\partial H/\partial R_i|n\rangle}{(E_n - E_m)^2} \ . \tag{2.16}$$

The Berry curvature can be understood in a more intuitive way from the above equation. It results from the virtual excitations driven by the time evolution of the Hamiltonian. In the adiabatic approximation, the system starting with a certain eigenstate must stay in the same instantaneous eigenstate. However, there could be virtual process that the system tunnels to a different eigenstate then hops back. The effect of such process is manifested as the Berry curvature, and enters into the phase of the wavefunction after a cyclic evolution.

The Berry connection and the Berry curvature are analogous to the gauge field $\mathbf{A}(\mathbf{r})$ and magnetic field $\mathbf{B}(\mathbf{r}) = \nabla \times \mathbf{A}(\mathbf{r})$ in real space. To be specific, for a three dimensional parameter space $\mathbf{R} = (R_1, R_2, R_3)$, Eq. (2.13) can be simplified as

$$\boldsymbol{\Omega} = \nabla_{\mathbf{R}} \times \mathbf{A}(\mathbf{R}) \ , \tag{2.17}$$

where the ith component of the vector $\mathbf{\Omega}$ is defined as

$$\Omega_i = \epsilon_{ijk}\Omega_{jk} \ . \tag{2.18}$$

Here Ω_{jk} is the Berry curvature defined in Eq. (2.13) and ϵ_{ijk} is the Levi-Civita symbol. Therefore, the Berry curvature can be considered as the "magnetic field" in the parameter space $\mathbf{R}$, with $\mathbf{A}(\mathbf{R})$ being the corresponding gauge field. More interestingly, from Eq. (2.16), it is evident that the Berry curvature diverges if there is a degeneracy in the energy levels at some point $\mathbf{R}_0$ in the parameter space. It means a degeneracy point acts as a source that generates Berry curvature in the parameter space, and can be considered as a "magnetic monopole" in $\mathbf{R}$ space.

In the following, we will introduce the Berry phase of Bloch states in the BZ, which was first discussed by Zak in 1989 [13].

2.1.1 Berry phase of Bloch bands

As introduced in Chapter 1, the eigenstates of a single-body effective Hamiltonian in a periodic potential obey the Bloch theorem

$$|\psi_{n\mathbf{k}}\rangle = e^{i\mathbf{k}\cdot\mathbf{r}}|u_{n\mathbf{k}}\rangle \ , \tag{2.19}$$

where $|u_{n\mathbf{k}}\rangle$ is periodic in units of lattice vectors and $\mathbf{k}$ is the wavevector defined in reciprocal space. As a result of real-space periodicity, the reciprocal space is also periodic in units of reciprocal space lattice vectors $\mathbf{G}_i$, with $i = 1, ..., d$ for $d-$dimensional lattices. People usually define the Wigner-Seitz cell in reciprocal space as the first Brillouin zone, or simply as the Brillouin zone (BZ). If we adopt a "periodic gauge" for the Bloch states in the BZ, i.e., $|\psi_{n\mathbf{k}}\rangle = |\psi_{n\mathbf{k}+\mathbf{G}_i}\rangle$, then the BZ naturally acts as a cyclic parameter space for the Bloch states.

Following the discussion by Zak [13], let us now derive an expression for the Berry phase of Bloch states in crystalline solids. Imagine that a weak time-dependent electric

field with low frequency is applied to a 1D crystal. The single-particle Hamiltonian is then

$$H(t) = \frac{1}{2m}\Big(-i\hbar\partial_x - eA(t)\Big)^2 + V(x) \tag{2.20}$$

where $V(x)$ is periodic in units of lattice constant a, and we choose the gauge $A(t) = -E_0 t$, with E_0 to be the strength of the weak electric field. For an electric field that varies slowly as time, we can assume that there are a set of orthonormal instantaneous eigenfunctions at each time t for $H(t)$

$$H(t)\,\psi_{nk(t)} = \epsilon_{nk(t)}\,\psi_{nk(t)}\ , \tag{2.21}$$

where

$$\psi_{nk(t)} = e^{ikx} u_{nk(t)}\ . \tag{2.22}$$

As shown in the above equation, the time dependence of $\psi_{nk(t)}$ enters only through the wavevector $k(t) = k - eA(t)/\hbar$; and it is assumed that the wavevector of the plane wave is time-independent. We seek for a solution of the form of Eq. (2.22) for the Schrödinger equation given in Eq. (2.21).

Moreover, plugging Eq. (2.22) into Eq. (2.21), one obtains the Schrödinger equation for the periodic part of the Bloch state $u_{nk(t)}$

$$H_k(t)\,u_{nk(t)} = \epsilon_{nk(t)}\,u_{nk(t)}\ , \tag{2.23}$$

where

$$\begin{aligned} H_k(t) &= e^{-ikx} H(t) e^{ikx} \\ &= \frac{1}{2m}\Big(-i\hbar\partial_x + \hbar k - eA(t)\Big)^2 + V(x)\ . \end{aligned} \tag{2.24}$$

In the adiabatic approximation, as time increases from 0 to t, an electron starting in one of the Bloch states would remain in the corresponding instantaneous eigenstate, but the wavevector $k(t)$ varies in the BZ. Once $k(t)$ hits the boundary of the BZ, the

time-evolved wavefunction $\psi(t)$ returns to the original eigenstate at $k(0)$ and picks up an extra Berry phase. If it is not a periodic cycle, say, at some arbitrary time t,

$$\psi(t) = e^{i\beta_n(t)} e^{-\frac{i}{\hbar}\int_0^t \epsilon_{nk(t')}dt'} \psi_{nk(t)} \ , \tag{2.25}$$

Plugging Eq. (2.25) into the time-dependent Schrödinger equation, and using Eq. (2.23) and Eq. (2.24), one obtains the following equation for $\beta_n(t)$

$$\partial_t \beta_n(t) = i \langle u_{nk(t)} | \partial_t u_{nk(t)} \rangle \ , \tag{2.26}$$

where $\partial_t \equiv \partial/\partial t$. It means for a adiabatic cycle with period τ, the Berry phase $\beta_n(\tau)$ has the expression

$$\begin{aligned} \beta_n(\tau) &= \int_0^\tau dt\, i \langle u_{nk(t)} | \partial_t u_{nk(t)} \rangle \\ &= \int_{-\pi/a}^{\pi/a} dk\, i \langle u_{nk} | \partial_k u_{nk} \rangle \ . \end{aligned} \tag{2.27}$$

Eq. (2.27) can be easily generalized to 2D and 3D lattices. In 1D, a loop is formed in BZ only if one traverses throughout the BZ. In 2D and 3D, however, one can take an arbitrary loop in the BZ and calculate the corresponding Berry phase. The general expression is then

$$\beta_n(\mathcal{C}) = \oint_{\mathcal{C}} d\mathbf{k}\, i \langle u_{n\mathbf{k}} | \nabla_{\mathbf{k}} u_{n\mathbf{k}} \rangle \ . \tag{2.28}$$

2.1.2 Formulation in Wannier basis

We would like to show that the Berry phase of Bloch bands is closely related to the center of the corresponding Wannier functions. For simplicity, let us again consider a 1D lattice. The Wannier function $w_{nR_j}(x)$ associated with the nth Bloch band is

$$w_{nR_j}(x) = \frac{1}{\sqrt{N_s}} \sum_k \psi_{nk}(x) e^{-ikR_j} \ , \tag{2.29}$$

where N_s is the total number of cells in the lattice, and $N_s \to \infty$ in the thermodynamic limit, and $R_j = ja$ is a lattice vector with a being the lattice constant and j being an

integer. Conversely,

$$\psi_{nk}(x) = \frac{1}{\sqrt{N_s}} \sum_j e^{ikR_j} w_{nR_j}(x) \ . \tag{2.30}$$

The Berry phase of Bloch bands can be re-expressed in Wannier basis as

$$\begin{aligned} \beta_n &= i \int_{-\pi/a}^{\pi/a} dk \langle u_{nk} | \partial_k u_{nk} \rangle \\ &= \frac{2\pi}{a} \frac{1}{N_s^2} \sum_{j,j',k} e^{ik(R_j - R'_j)} \langle w_{nj} | x | w_{nj} \rangle \\ &= \frac{2\pi}{a} \frac{1}{N_s} \sum_j \langle w_{nj} | x | w_{nj} \rangle \\ &= 2\pi \frac{x_n}{a} \ , \end{aligned} \tag{2.31}$$

where $x_n = \langle w_{n0} | x | w_{n0} \rangle$ is the center of the Wannier function in the "home" unitcell, i.e., the $j = 0$ cell, which is also denoted as the "Wannier center" or "Wannier charge center" in the literatures [52, 53]. The substitution $(a/2\pi) \int_{-\pi/a}^{\pi/a} dk \to (1/N_s) \sum_k$ is used when going from the first to the second line. The Wannier-function formulation of Berry phase in Eq. (2.31) provides an intuitive interpretation to the Berry-phase theory of electric polarization [15, 54], which will be discussed later.

2.1.3 Generalized gauge freedom and non-Abelian Berry phase

In the above discussions, we have only considered the case of a single eigenstate which remains isolated from other eigenstates during the cyclic adiabatic evolution. However, in general there may be crossing of states during an adiabatic evolution, then the single-state formulation of Berry phase breaks down. For example, in solids, as a result of crystalline symmetry, the Bloch states at a wavevector $\mathbf{k}$ can be characterized by irreducible representations of the small symmetry group at $\mathbf{k}$. The degeneracy of the Bloch states at $\mathbf{k}$ is exactly the dimension of the corresponding irreducible representation (irreps) [55]. Therefore, typically there is high degeneracy at high-symmetry points in the BZ. Even for non-degenerate states that belong to one-dimensional irreducible representations, two energy levels may still touch each other without avoided crossing

as long as their symmetry characters are different. Besides crystalline symmetries, for fermionic systems with time-reversal symmetry, the Bloch states are guaranteed to be two-fold degenerate at time-reversal invariant momenta (TRIM) in the BZ [1]. Therefore, degeneracy is ubiquitous in crystalline solids, which means the single-band formulation for Berry phases (Eq. (2.27)) has to be generalized to fit multi-band situations.

In solids, usually there exists a group of energy bands which are separated from other bands by a finite energy gap at every $\mathbf{k}$ point in the BZ. For example, for insulators, the occupied (valence) bands are separated from the unoccupied (conduction) bands by an energy gap. In such situations, it is helpful to define the Berry connection, Berry curvature and Berry phase in terms of the entire set of bands, because in general degeneracy is unavoidable within such a group of bands. It is then natural to generalize the Berry connection and Berry curvature to multi-band cases [52], or, in other words, to "non-Abelian" cases

$$\begin{aligned} A_{a,mn} &= i\langle u_{m\mathbf{k}}|\partial_a u_{n\mathbf{k}}\rangle \ , \\ \Omega_{ab,mn} &= \partial_a A_{b,mn} - \partial_b A_{a,mn} \ , \end{aligned} \tag{2.32}$$

where $m, n = 1, ...N$ are band indices, and a, b label the $\mathbf{k}$-space directions, i.e., $\partial_a \equiv \partial/\partial_{k_a}$. Now both $A_{mn,a}$ and $\Omega_{mn,ab}$ are matrix elements of $N \times N$ matrices.

As discussed in Sec. 2.1, there is a $U(1)$ gauge freedom for electronic wavefunctions as shown in Eq. (2.10) (the case of superconductors is not considered in this thesis), which means a physical observable remains invariant against a local twist of phase. In insulating solids, as mentioned above, there is always a group of occupied energy bands which are separated from other bands by a finite energy gap, while degeneracy is generally present within this set of bands. In such a situation, a physical observable is usually calculated by tracing over the entire group of bands, and the matrix element

[1] TRIM means those wavevectors that are invariant under time-reversal operation modulo a reciprocal lattice vector. For example, in 1D, there are only two TRIM, $k = 0$ and $k = \pi$. In 2D, there are four TRIM; in 3D there are eight TRIM. The TRIM play a crucial role in theory of topological insulators which will be discussed in more details in Sec. 2.3.

of each individual band is not quite meaningful due to the entanglement among the bands. Therefore, in such multi-band (N-band) case, the $U(1)$ gauge freedom can be generalized to a $U(N)$ gauge freedom. A physical observable should be invariant under the following $U(N)$ transformation

$$|u_{m\mathbf{k}}\rangle \to \sum_{m=1}^{N} |u_{m\mathbf{k}}\rangle\, U_{mn} \ , \tag{2.33}$$

where U_{mn} is the matrix element of a $N \times N$ unitary matrix.

One can check that the non-Abelian Berry connection and Berry curvature defined in Eq. (2.32) are not "gauge-covariant" under a $U(N)$ transformation. Instead, they transform as

$$\begin{aligned} A_a &\to U^\dagger\, A_a\, U + i\, U^\dagger\, \partial_a U \ , \\ \Omega_{ab} &\to U^\dagger\, \Omega_{ab}\, U + \Lambda_{ab} + i\, [C_a, U^\dagger A_b U] - i\, [C_b, U^\dagger A_a U] \ , \end{aligned} \tag{2.34}$$

where

$$\begin{aligned} C_a &= i\, U^\dagger\, \partial_a U \ , \\ \Lambda_{ab} &= \partial_a C_b - \partial_b C_a \ , \end{aligned} \tag{2.35}$$

can be considered as the "Berry connection" and "Berry curvature" in the gauge space.

Since the Berry curvature matrix now is no longer "gauge-covariant", we can define a new Berry-curvature-like quantity which transforms covariantly under the gauge transformation,

$$\widetilde{\Omega}_{ab} = \Omega_{ab} - i\, [A_a, A_b] \ . \tag{2.36}$$

It is easy to check that $\widetilde{\Omega}_{ab}$ is gauge-covariant, i.e., under a $U(N)$ transformation, $\widetilde{\Omega}_{ab} \to U^\dagger \widetilde{\Omega}_{ab} U$.

One can still calculate the Berry phase in such non-Abelian cases, but now the loop integration of the Berry connection gives us a matrix instead of a scalar. Moreover, this "Berry-phase" matrix in general is non-diagonal. What we can do is to diagonalize the Berry-phase matrix, and consider each eigenvalue as the Berry phase of the

corresponding unitarily transformed Bloch state. However, the Berry phase of each individual band is no longer gauge invariant. Only the sum of them is a gauge-invariant quantity modulo 2π.

2.1.4 Symmetry properties

The presence of symmetries, such as time-reversal ($\mathcal{T}$) and inversion ($\mathcal{P}$) symmetries, may impose strong constraints on possible values of Berry phase. In this section, we will particularly discuss systems with $\mathcal{T}$ symmetry only and with both $\mathcal{T}$ and $\mathcal{P}$ symmetries.

Time-reversal symmetry

For a lattice with $\mathcal{T}$ symmetry, one can insist on a gauge choice for the Bloch states that respect $\mathcal{T}$ symmetry (which cannot be both smooth and periodic for $\mathbb{Z}_2$ topological insulators [56, 57]). We have to distinguish two situations: (a) spinless electrons, such as electrons in graphene and other nonmagnetic materials with negligible SOC; (b) spin-1/2 fermions, such as electrons in materials with non-negligible spin-orbit coupling. Since there is $\mathcal{T}$ symmetry, a magnetic order is allowed in neither case.

A gauge respecting $\mathcal{T}$ symmetry means that the Bloch function at $\mathbf{k}$ is related to the one at $-\mathbf{k}$ by a time-reversal operation without any gauge ambiguity. For spinless electrons, the time-reversal operation is simply a complex-conjugation operation, which means

$$\begin{aligned} \psi_{n\mathbf{k}}(\mathbf{r}) &= \psi^*_{n-\mathbf{k}}(\mathbf{r}) \ , \\ u_{n\mathbf{k}}(\mathbf{r}) &= u^*_{n-\mathbf{k}}(\mathbf{r}) \ . \end{aligned} \tag{2.37}$$

The above gauge choice imposes the following constriant on the Berry connection matrix

$$\begin{aligned} A_{a,mn}(\mathbf{k}) &= i\,\langle u_{m\mathbf{k}}|\partial_a u_{n\mathbf{k}}\rangle \\ &= i\int d\mathbf{r}\, u^*_{m\mathbf{k}}(\mathbf{r})\,\partial_a u_{n\mathbf{k}}(\mathbf{r}) \\ &= i\int d\mathbf{r}\, u_{m-\mathbf{k}}(\mathbf{r})\,\partial_a u^*_{n-\mathbf{k}}(\mathbf{r}) \\ &= i\int d\mathbf{r}\, u^*_{n-\mathbf{k}}(\mathbf{r})\,\partial_{-a} u_{m-\mathbf{k}}(\mathbf{r}) \\ &= A_{a,nm}(-\mathbf{k})\ , \end{aligned} \tag{2.38}$$

where $\partial_{-a} \equiv \partial_{-k_a}$ in the above equation. Using similar tricks, it is straightforward to show that the Berry curvature and the gauge-covariant Berry curvature matrix (Eq. (2.36)) behaves as odd functions of $\mathbf{k}$:

$$\begin{aligned} \Omega_{ab,mn}(\mathbf{k}) &= -\Omega_{ab,nm}(-\mathbf{k})\ , \\ \widetilde{\Omega}_{ab,mn}(\mathbf{k}) &= -\widetilde{\Omega}_{ab,nm}(-\mathbf{k})\ . \end{aligned} \tag{2.39}$$

Writing Eq. (2.38) and Eq. (2.39) in matrix form in the basis of the occupied Bloch bands, one obtains

$$A_a(\mathbf{k}) = A_a^T(-\mathbf{k})\ , \tag{2.40}$$

$$\Omega_{ab}(\mathbf{k}) = -\Omega_{ab}^T(-\mathbf{k})\ , \tag{2.41}$$

$$\widetilde{\Omega}_{ab}(\mathbf{k}) = -\widetilde{\Omega}_{ab}^T(-\mathbf{k})\ , \tag{2.42}$$

where the superscript "T" refers to matrix transpose.

For spin-1/2 fermionic systems, there must be an even number of bands (say, $2N$ bands). In most cases one can decompose the $2N$ bands into N pairs. Kramers' degeneracy is satisfied in each pair of bands

$$\begin{aligned} \epsilon_{n\mathbf{k},1} &= \epsilon_{n-\mathbf{k},2}\ , \\ \epsilon_{n\mathbf{k},2} &= \epsilon_{n-\mathbf{k},1}\ , \end{aligned} \tag{2.43}$$

where the subindices "1" and "2" refer to the spin indices in each pair [2]. The gauge is

[2]Here we refer to the local spin polarization at each $\mathbf{k}$, whose direction may vary with $\mathbf{k}$.

said to respect $\mathcal{T}$ symmetry if for the nth pair ($1 \leq n \leq N$), the following relationship is satisfied

$$\begin{aligned} \mathcal{T}|u_{n\mathbf{k},1}\rangle &= -|u_{n-\mathbf{k},2}\rangle \\ \mathcal{T}|u_{n\mathbf{k},2}\rangle &= |u_{n-\mathbf{k},1}\rangle \end{aligned} \tag{2.44}$$

Now $\mathcal{T} = i\sigma_y K$, where σ_y is the second Pauli matrix acting on the spin subspace within each Kramers pair, and K is the complex-conjugation operator. More specifically, in real-space representation,

$$\begin{aligned} u_{n\mathbf{k},1}(\mathbf{r}) &= -u^*_{n-\mathbf{k},2}(\mathbf{r}) \ , \\ u_{n\mathbf{k},2}(\mathbf{r}) &= u^*_{n-\mathbf{k},1}(\mathbf{r}) \ . \end{aligned} \tag{2.45}$$

As a result of Eq. (2.45), the matrix elements of the non-Abelian Berry connection obey the following relationship

$$\begin{aligned} A_{a,mn,11}(\mathbf{k}) &= i\langle u_{m\mathbf{k},1}|\partial_a u_{n\mathbf{k},1}\rangle \\ &= i\int d\mathbf{r}\, u^*_{m\mathbf{k},1}(\mathbf{r})\, \partial_a u_{n\mathbf{k},1}(\mathbf{r}) \\ &= i\int d\mathbf{r}\, u_{m-\mathbf{k},2}(\mathbf{r})\, \partial_a u^*_{n-\mathbf{k},2}(\mathbf{r}) \\ &= i\int d\mathbf{r}\, u^*_{n-\mathbf{k},2}(\mathbf{r})\, \partial_{-a} u_{m-\mathbf{k},2}(\mathbf{r}) \\ &= A_{a,nm,22}(-\mathbf{k}) \ , \end{aligned} \tag{2.46}$$

and

$$\begin{aligned} A_{a,mn,12}(\mathbf{k}) &= i\langle u_{m\mathbf{k},1}|\partial_a u_{n\mathbf{k},2}\rangle \\ &= i\int d\mathbf{r}\, u^*_{m\mathbf{k},1}(\mathbf{r})\, \partial_a u_{n\mathbf{k},2}(\mathbf{r}) \\ &= -i\int d\mathbf{r}\, u_{m-\mathbf{k},2}(\mathbf{r})\, \partial_a u^*_{n-\mathbf{k},1}(\mathbf{r}) \\ &= -i\int d\mathbf{r}\, u^*_{n-\mathbf{k},1}(\mathbf{r})\, \partial_{-a} u_{m-\mathbf{k},2}(\mathbf{r}) \\ &= -A_{a,nm,12}(-\mathbf{k}) \ . \end{aligned} \tag{2.47}$$

Similarly,

$$A_{a,mn,21}(\mathbf{k}) = -A_{a,nm,21}(-\mathbf{k})$$
$$A_{a,mn,22}(\mathbf{k}) = A_{a,nm,11}(-\mathbf{k}) \ . \tag{2.48}$$

The above equations can be written in a neat matrix form

$$A_a(\mathbf{k}) = \sigma_y \, A_a^T(-\mathbf{k}) \, \sigma_y \ , \tag{2.49}$$

where σ_y acts only on the (pseudo)spin subspace in each Kramers' pair, while the transpose operation "T" acts on the entire space of the $2N$ bands. A similar relationship can be obtained for the non-Abelian Berry curvature,

$$\Omega_{ab}(\mathbf{k}) = -\sigma_y \, \Omega_{ab}^T(-\mathbf{k}) \, \sigma_y \ . \tag{2.50}$$

It should be noted that Eq. (2.40)-(2.42) and Eq. (2.49)-(2.50) are valid only for the time-reversal invariant gauge as shown in Eq. (2.37) and Eq. (2.45).

Now we are ready to discuss the constriants on the Berry phase from $\mathcal{T}$ symmetry. For spinless electrons in 2D and 3D lattices [3], if a $\mathbf{k}$ loop is chosen in such a way that for any $\mathbf{k}$ on the loop, $-\mathbf{k}$ is also on the loop (we call such a loop a $\mathcal{T}$-symmetric loop), as shown in Fig. 2.1(b), then the total Berry phase for a set of isolated bands around this loop must be vanishing, i.e.,

$$\begin{aligned} \beta &= \oint_{\mathcal{C}_{\rm TR}} d\mathbf{k} \cdot \, {\rm tr} \, [\mathbf{A}(\mathbf{k})] \\ &= \int_{\mathcal{S}} dk_a dk_b \, {\rm tr} \, [\Omega_{ab}(\mathbf{k})] = 0 \end{aligned} \tag{2.51}$$

where $\mathcal{C}_{\rm TR}$ is the "$\mathcal{T}$-symmetric loop" as described above (Fig. 2.1(b)), $\mathcal{S}$ is the area enclosed by the loop $\mathcal{C}_{\rm TR}$, and "tr" refers to the trace over the energy bands. From Eq. (2.41), we know that the trace of the Berry curvature at $\mathbf{k}$ and $-\mathbf{k}$ exactly cancel each other, leading to a vanishing total Berry phase.

[3] In 3D systems, we have to consider time-reversal-invariant $\mathbf{k}$ planes, e.g., the 2D planes at $k_j = 0, \pi$.

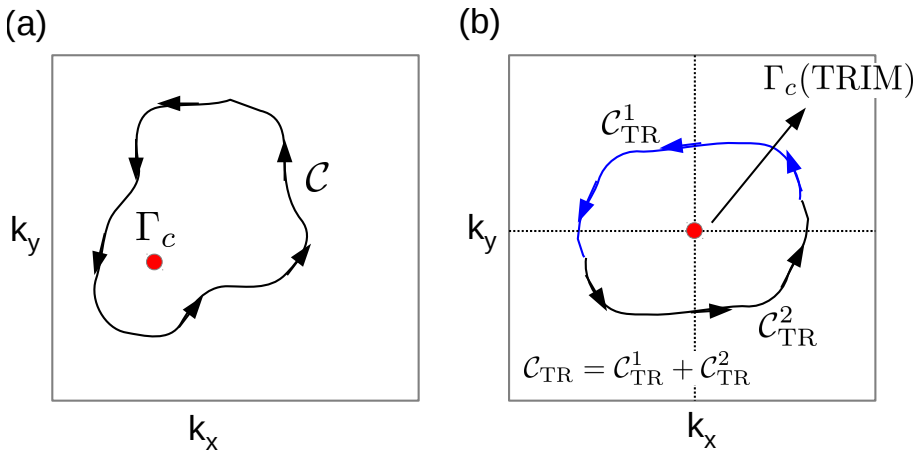

Figure 2.1: Schematic plots of the $\mathbf{k}$ loops. (a) An arbitrary loop $\mathcal{C}$ in the BZ enclosing a degeneracy point Γ_c. (b) A $\mathcal{T}$-symmetric loop $\mathcal{C}_{\rm TR}$ which is consisted of two paths $\mathcal{C}^1_{\rm TR}$ (blue) and $\mathcal{C}^2_{\rm TR}$ (black). $\mathcal{C}^2_{\rm TR}$ can be obtained from $\mathcal{C}^1_{\rm TR}$ by a $\mathcal{T}$ operation, i.e., $\mathbf{k} \to -\mathbf{k}$, and the choice of $\mathcal{C}^1_{\rm TR}$ is arbitrary. A degeneracy point Γ_c resided at one of the TRIM is enclosed by the loop $\mathcal{C}_{\rm TR}$.

It may also be interesting to look at the Berry phase of each individual band. One can diagonalize the Berry connection matrix $A_a(\mathbf{k}) = V^\dagger(\mathbf{k})\, A^{\rm D}_a(\mathbf{k})\, V(\mathbf{k})$, where $V(\mathbf{k})$ is the eigenvector matrix of $A_a(\mathbf{k})$, and $A^{\rm D}_a(\mathbf{k})$ is the corresponding diagonal eigenvalue matrix. The integration of each eigenvalue $A^{\rm D}_{a,nn}(\mathbf{k})$ around the loop $\mathcal{C}_{\rm TR}$ would be the Berry phase of the nth Bloch-like state, which is unitarily transformed from the original gauge by $V(\mathbf{k})$, $|u_{n\mathbf{k}}\rangle = \sum_m |u^0_{m\mathbf{k}}\rangle\, V_{mn}(\mathbf{k})$, where $|u^0_{m\mathbf{k}}\rangle$ refers to the original gauge. Let us denote such a single-band Berry phase as β_n,

$$\beta_n = \oint_{\mathcal{C}_{\rm TR}} d\mathbf{k} \cdot \mathbf{A}^{\rm D}_{nn}(\mathbf{k}) \; . \tag{2.52}$$

From Eq. (2.39), it is plausible to say that since $A^{\rm D}_{a,nn}(\mathbf{k}) = A^{\rm D}_{a,nn}(-\mathbf{k})$, β_n has to vanish as a result of the exact cancellation between two opposite $\mathbf{k}$ points. However, this is not necessarily true if there is a degeneracy at the center of $\mathcal{S}$ enclosed by $\mathcal{C}_{\rm TR}$ (denoted as Γ_c), which by definition has to be one of the TRIM. Such a degeneracy is quite possible at high-symmetry points for a group of entangled bands. From (2.16), we know that the Berry curvature for band n would diverge at Γ_c as a result of the degeneracy. Moreover, as Γ_c is invariant under $\mathcal{T}$ operation, there is no way to cancel the diverging Berry curvature.

On the other hand, from Eq. (2.52) we know that if $\mathbf{k} \to -\mathbf{k}$, the Berry phase flips its sign $\beta_n \to -\beta_n$, which in turn must be equal to β_n as a result of $\mathcal{T}$ symmetry. Since β_n is gauge invariant modulo 2π, there are only two possible values for β_n, either 0 or π. If there is no degeneracy at the center Γ_c (nor anywhere else within the loop), we know that $\beta_n = 0$ due to the cancellation of Berry connections at opposite $\mathbf{k}$ points. It thus follows that when there is degeneracy at Γ_c, $\beta_n = \pi$.

To summarize, for spinless electronic systems with $\mathcal{T}$ symmetry only, the Berry phase of the nth "disentangled" band around a $\mathcal{T}$-symmetric loop centered at Γ_c can only take two possible values

$$\beta_n(\mathcal{C}_{\rm TR}) = \begin{cases} \pi & \text{if there is a degeneracy at } \Gamma_c \ , \\ 0 & \text{otherwise} \ . \end{cases} \tag{2.53}$$

The above conclusion for spinless particles can be generalized to spin-1/2 fermions. Again, let us consider the situation that there are N pairs of bands which are separated from other bands by a finite energy gap. The Kramers' degeneracy is satisfied within each pair (Eq. (2.43)), and a gauge respecting $\mathcal{T}$ symmetry has been constructed (Eq. (2.45)). For such a situation, it is helpful to divide a $\mathcal{T}$-symmetric loop $\mathcal{C}_{\rm TR}$ into two paths as shown in Fig. 2.1(b), $\mathcal{C}_{\rm TR} = \mathcal{C}^1_{\rm TR} + \mathcal{C}^2_{\rm TR}$, where $\mathcal{C}^2_{\rm TR}$ can be obtained from $\mathcal{C}^1_{\rm TR}$ by a time-reversal operation $\mathbf{k} \to -\mathbf{k}$, and the choice of $\mathcal{C}^1_{\rm TR}$ is indeed arbitrary. Then let us define the Berry phase $\beta_{n,h}$ and $\beta_{n,l}$ as the Berry phases of the Bloch states which are higher and lower in energy for the nth Kramers' pair,

$$\begin{aligned} \beta_{n,h}(\mathcal{C}_{\rm TR}) &= \int_{\mathcal{C}^1_{\rm TR}} d\mathbf{k} \cdot \ i \langle u_{n\mathbf{k},1} | \nabla_{\mathbf{k}} u_{n\mathbf{k},1} \rangle + \int_{\mathcal{C}^2_{\rm TR}} d\mathbf{k} \cdot \ i \langle u_{n\mathbf{k},2} | \nabla_{\mathbf{k}} u_{n\mathbf{k},2} \rangle \ , \\ \beta_{n,l}(\mathcal{C}_{\rm TR}) &= \int_{\mathcal{C}^1_{\rm TR}} d\mathbf{k} \cdot \ i \langle u_{n\mathbf{k},2} | \nabla_{\mathbf{k}} u_{n\mathbf{k},2} \rangle + \int_{\mathcal{C}^2_{\rm TR}} d\mathbf{k} \cdot \ i \langle u_{n\mathbf{k},1} | \nabla_{\mathbf{k}} u_{n\mathbf{k},1} \rangle \ . \end{aligned} \tag{2.54}$$

Due to Kramers' degeneracy, the Berry curvature of each individual band must diverge at the TRIM. As a result, the Berry phases $\beta_{n,h}(\mathcal{C}_{\rm TR})$ and $\beta_{n,l}(\mathcal{C}_{\rm TR})$ around a

$\mathcal{T}$ symmetric loop $\mathcal{C}_{\rm TR}$ enclosing one of the TRIM Γ_c have to be $\pm\pi$, i.e.,

$$\beta_{n,h}(\mathcal{C}_{\rm TR}) = \pm\pi \ ,$$
$$\beta_{n,l}(\mathcal{C}_{\rm TR}) = \mp\pi \ . \tag{2.55}$$

Eq. (2.55) explains the quantized (π) Berry phase in systems with strong Rashba spin-orbit coupling, which gives rise to the anti-weak localization in semiconductor heterostructures with the breaking of inversion symmetry [58]. From Eq. (2.55), it is also evident that the sum of the Berry phases in each Kramers' pair is still vanishing, i.e., $\beta_{n,h} + \beta_{n,l} = 0$.

Time-reversal and inversion symmetries

Now let us proceed to the situation with both $\mathcal{T}$ and $\mathcal{P}$ symmetries. Since the spatial inversion operation acts only on the orbital degrees of freedom, it is not necessary to distinguish between spin-1/2 and spinless particles. The inversion operation maps $u_{n\mathbf{k}}(\mathbf{r})$ to $u_{n-\mathbf{k}}(\mathbf{r})$,

$$\mathcal{P}u_{n\mathbf{k}}(\mathbf{r}) = u_{n-\mathbf{k}}(\mathbf{r}) = e^{i\phi(\mathbf{k})}u_{n\mathbf{k}}(\mathbf{r}) \ . \tag{2.56}$$

Then one can construct a gauge respecting $\mathcal{P}$ symmetry by fixing the otherwise arbitrary phase factor as 0, i.e., $\phi(\mathbf{k}) = 0$,

$$u_{n\mathbf{k}}(\mathbf{r}) = u_{n-\mathbf{k}}(\mathbf{r}) \ . \tag{2.57}$$

If the periodic part of the Bloch functions satisfy the above equation, we say a gauge respecting $\mathcal{P}$ symmetry has been constructed.

If both $\mathcal{T}$ and $\mathcal{P}$ symmetries are present, we have to distinguish between spinless electrons and spin-1/2 fermions. For spinless electrons, comparing Eq. (2.37) with Eq. (2.57), it is obvious that $u_{n\mathbf{k}}(\mathbf{r})$ can be made real everywhere in the BZ. It implies that the Berry phase of each individual band around any loop in the BZ must be either 0 or $\pm\pi$ in order to be compatible with the reality condition.

The above conclusion may be further illustrated by looking at the properties of the Berry curvature. For spinless systems with $\mathcal{PT}$ symmetry, it is straightforward to show that the Berry curvature and Berry connection matrices have to be skew-symmetric,

$$A_a(\mathbf{k}) = -A_a^T(\mathbf{k}) \ , \tag{2.58}$$

$$\Omega_{ab}(\mathbf{k}) = -\Omega_{ab}^T(\mathbf{k}) \ , \tag{2.59}$$

which means the trace of them vanish everywhere in the BZ. It follows that the total Berry phase for an isolated set of bands must vanish for any $\mathbf{k}$ loop in the BZ.

However, as discussed previously, if we are interested in the Berry phase of each individual band β_n, then β_n could be either 0 or $\pm\pi$. When a degeneracy point is enclosed by the $\mathbf{k}$ loop, $\beta_n = \pm\pi$ [59], otherwise it is 0. In other words, for spinless systems with both $\mathcal{T}$ and $\mathcal{P}$ symmetries,

$$\beta_n(\mathcal{C}) = \begin{cases} \pi & \text{if there is a degeneracy point enclosed by the loop } \mathcal{C} \\ 0 & \text{otherwise} \end{cases} \tag{2.60}$$

The difference between Eq. (2.53) and Eq. (2.60) deserves discussion. In the former case with $\mathcal{T}$ symmetry only, the quantization condition of the Berry phase applies only if the loop is chosen as a $\mathcal{T}$-symmetric loop $\mathcal{C}_{\mathrm{TR}}$ centered at one of the TRIM as shown in Fig. 2.1(b). However, the conclusion shown in Eq. (2.53) is valid for both spinless and spin-1/2 systems. On the other hand, in the latter case with both $\mathcal{T}$ and $\mathcal{P}$ symmetries, the loop can be an arbitrary loop $\mathcal{C}$ in the BZ and the degeneracy point can be anywhere within the loop as shown in Fig. 2.1(a), but the quantization condition (Eq. (2.60)) applies only to spinless systems. The π Berry phase around the Dirac cone in graphene [60] can be explained by Eq. (2.60).

The case of spin-1/2 fermions is a little bit subtle. As mentioned above, the $\mathcal{P}$ symmetry does not distinguish the spin, thus the only difference would be from the $\mathcal{T}$ symmetry. A $\mathcal{PT}$ operation would map the Bloch state at $\mathbf{k}$ with up (pseudo) spin to

the state with down (pseudo) spin at the same $\mathbf{k}$ point,

$$u_{n\mathbf{k},1}(\mathbf{r}) = e^{i\phi(\mathbf{k})} u^*_{n\mathbf{k},2}(\mathbf{r}) \ ,$$
$$u_{n\mathbf{k},2}(\mathbf{r}) = -e^{-i\phi(\mathbf{k})} u^*_{n\mathbf{k},1}(\mathbf{r}) \ , \tag{2.61}$$

where $\phi(\mathbf{k})$ is an arbitrary phase factor. As a result, the Bloch bands are two-fold degenerate everywhere in the BZ,

$$\epsilon_{n\mathbf{k},1} = \epsilon_{n\mathbf{k},2} \ , \tag{2.62}$$

where n is the index of the Kramers' pair, while "1" and "2" are the spin indices.

Let us consider the situation that the Kramers' pairs are separated from each other by finite energy gaps. Then the Berry connection and the Berry curvature for the nth Kramers' pair are 2×2 matrices due to the degeneracy. Setting $\phi(\mathbf{k}) = 0$, one would find that the Berry connection matrix elements for the nth Kramers' pair obey the following relationship,

$$\mathbf{A}_{n,11}(\mathbf{k}) = -\mathbf{A}_{n,22}(\mathbf{k}) \ . \tag{2.63}$$

It immediately follows from Eq. (2.63) that the total Berry phase around any loop in the BZ must be vanishing. If one locally diagonalizes the Berry connection matrix $\mathbf{A}_n(\mathbf{k})$ and calculate the Berry phases of the two "disentangled" subbands for each Kramers' pair, one would find that the two Berry phases are equal and opposite. If a $\mathcal{T}$-symmetric loop (Fig. 2.1(b)) is chosen in the BZ, then according to Eq. (2.53), $\beta_{n,h}$ and $\beta_{n,l}$ as defined in Eq. (2.54) have to be quantized as $\pm\pi$, i.e., $\beta_{n,h} = \pm\pi$, and $\beta_{n,l} = \mp\pi$.

In summary, in this section we have introduced the concept of Berry phase and discussed some properties of the Berry phase. In the following section, we will start talking about the applications of Berry phase in solid state physics.

2.2 Examples of Berry phase

In the above section we have introduced the Berry phase. In this section, we will focus on the applications of Berry phase in specific physical problems. We start with the simplest example, a two-level system, where the Berry connection and the Berry curvature have analytic expressions. Then we discuss the electric polarizations in one-dimensional systems, the anomalous Hall conductance (conductivity) in 2D (3D) systems, and the Chern-Simons orbital magnetoelectric coupling in 3D systems. All three physical quantities can be regarded as geometric properties of the occupied Bloch functions.

2.2.1 Two-level system

The two-level system may be one of the most "classical" examples to illustrate the properties of the Berry phase. It has a simple analytic solution, and it provides a natural connection between the geometric phase and the topological invariant. More importantly, it may be considered as a two-band effective Hamiltonian from a $\mathbf{k}\cdot\mathbf{p}$ model expanded around a given high-symmetry point in the BZ, which is widely used in the band theory [12].

We consider a spin or a pseudospin σ, which couples to a "Zeeman field" parameterized by $\mathbf{R}$

$$H = \sigma \cdot \mathbf{h}(\mathbf{R}) \ . \tag{2.64}$$

Here σ refer to the Pauli matrices representing the pseudospin. Let us consider the case that $\mathbf{h}$ is defined on the surface of a unit sphere, which has a constant magnitude and points outward from the center of the sphere

$$\mathbf{h}(\mathbf{R}) = h\hat{\mathbf{R}} = h(\sin\theta\cos\phi, \sin\theta\cos\phi, \cos\theta) \ , \tag{2.65}$$

where $\theta \in [0, \pi]$ and $\phi \in [0, 2\pi]$ are introduced as the polar and azimuthal angles in the spherical coordinate system.

Following Ref. [12], if one fixes the gauge for one of the eigenstates (ψ_1) at the north pole ($\theta = 0$), then ψ_1 can be chosen to be smooth and single-valued everywhere on the sphere except for the south pole ($\theta = \pi$):

$$|\psi_1\rangle = (\sin\frac{\theta}{2}e^{-i\phi}, -\cos\frac{\theta}{2}) ,$$
$$|\psi_2\rangle = (\cos\frac{\theta}{2}e^{-i\phi}, \sin\frac{\theta}{2}) , \tag{2.66}$$

where $|\psi_1\rangle$ ($|\psi_2\rangle$) is the eigenstate with the negative (positive) eigenvalue. With such a gauge choice, $|\psi_1\rangle$ at the south pole are dependent on the azimuthal angle ϕ, implying that they are multivalued at this point. One can certainly make another gauge choice, for example, $|\psi_1\rangle \to |\psi_1\rangle e^{i\phi}$, such that the states are single-valued at the south pole, but it turns out to be multivalued at the north pole [12, 61]. Actually it is impossible to insist a smooth and single-valued gauge for either one of the two eigenstates over the entire sphere. Some discontinuity or multivaludness must show up somewhere on the sphere. The best one can do is to push it to a point such as the south pole or the north pole. As will be discussed later in this section, this is due to a topological obstruction.

Taking the gauge shown in Eq. (2.66), the Berry curvature of $|\psi_1\rangle$ turns out to be

$$\Omega_{\theta\phi} = \frac{1}{2}\sin\theta , \tag{2.67}$$

which is gauge invariant. The corresponding Berry phase around an arbitrary loop $\mathcal{C}$ on the surface of the sphere is then

$$\beta(\mathcal{C}) = \frac{1}{2}\int_{\mathcal{S}} \sin\theta d\theta d\phi = \frac{1}{2}\Lambda(\mathcal{S}) , \tag{2.68}$$

where $\mathcal{S}$ is the area enclosed by loop $\mathcal{C}$, and $\Lambda(\mathcal{S})$ is the solid angle subtended by $\mathcal{S}$. If $\mathcal{S}$ is chosen as the entire surface of the sphere, it follows from Eq. (2.68) that the Berry phase is quantized as 2π, i.e., $\beta_0 = 2\pi C$, where $C = 1$ in our case. It turns out that the integer C is a topological invariant, which is known as the Chern number of $|\Psi_1\rangle$ defined on the surface of the sphere. It can be an arbitrary integer in more general cases.

The topological nature of the model becomes more explicit if we extend the parameter space to the entire 3D Euclidean space, instead of on the surface of the sphere. In other words, the magnitude of $\mathbf{h}$ is considered as the third variable. In the spherical coordinate system, the Berry connections of $|\psi_1\rangle$ are given by the following expressions,

$$\begin{aligned} A_r &= i\langle\psi_1|\frac{\partial}{\partial h}|\psi_1\rangle = 0 \ , \\ A_\theta &= i\langle\psi_1|\frac{\partial}{h\partial\theta}|\psi_1\rangle = 0 \ , \\ A_\phi &= i\langle\psi_1|\frac{\partial}{h\sin\theta\partial\phi}|\psi_1\rangle = \frac{1-\cos\theta}{2h\sin\theta} \ . \end{aligned} \tag{2.69}$$

The Berry curvature is a 3-vector in 3D parameter-space, which is simply the curl of the Berry connection (Eq. (2.17)). Apply the curl operator in the spherical coordinate system to the Berry connections defined in Eq. (2.69), one obtains the expression of the Berry-curvature vector (Eq. (2.18))

$$\mathbf{\Omega} = \frac{1}{2h^2}\hat{\mathbf{R}} \ , \tag{2.70}$$

where $\hat{\mathbf{R}}$ is the unit vector pointing to the radial direction. From Eq. (2.70), we see that the expression of $\mathbf{\Omega}$ strongly resembles the magnetic field generated by a magnetic monopole located at the center of the sphere. It is indeed helpful to interpret the degeneracy point at the origin as a "magnetic monopole" in the parameter space. Then the Berry curvature $\underline{0}$ may be regarded as "magnetic field" generated by such a monopole. Gauss's theorem tells us that for any closed surface enclosing the monopole, the surface integral of the Berry curvature must be quantized and proportional to the number of monopoles enclosed by the surface. This is why the Berry phase over the entire surface of the sphere must be quantized as $2\pi C$, where C is exactly the number of monopoles enclosed by the surface. It is the non-zero Chern number that brings the obstruction to constructing a smooth and single-valued (or, smooth and periodic) gauge over the entire surface of the sphere [62].

2.2.2 Electric polarization

Electric polarization is one of the most fundamental physical quantities in solids. In almost all the traditional textbooks about electromagnetism, the electric polarization is defined as some kind of electric-dipole density. In crystalline solids, however, the polarization may be partitioned into two parts: one contributed from the bare nuclei and the localized core electrons, denoted as $\mathbf{P}_{\text{ion}}$, and the other from the valence electrons denoted as $\mathbf{P}_{\text{elec}}$. The bare nuclei are screened by the core electrons, thus the bare nuclear charges are usually renormalized as some effective nuclear charges. In crystalline systems, the contribution to the total polarization $\mathbf{P}$ from such kind of screened nuclei is simply the sum of all the nuclear dipole moments in a primitive cell divided by the volume of the cell

$$\mathbf{P}_{\text{ion}} = \frac{1}{V_{\text{cell}}} \sum_i Z_i \, \mathbf{R}_i \ , \tag{2.71}$$

where Z_i is the effective nuclear charge of the ith ion and $\mathbf{R}_i$ is the ionic position. Such a definition is valid for the nuclear contribution because the core electrons are deeply localized, and are tightly bound with the nuclei. Moreover, the nuclei and the core electrons bound with them are well separated from each other in space, which means they can be treated as some kind of classical objects.

One may attempt to define the contribution from valence electrons in a similar approach. If all the valence electrons are localized around the nuclei such that there is vanishing interstitial charge distribution, then Eq. (2.71) still works, except that Z_i is interpreted as the ionic charge including the valence electrons. The above definition may be a good approximation for ionic lattices such as NaCl, but it clearly fails for materials with covalent bonds where the valence charge density extends through the entire lattice [54]. In the latter case, it is unclear how to unambiguously define the ionic charge.

Another attempt to define $\mathbf{P}_{\text{elec}}$ is to generalize the discrete summation of the localized charge to an integration of the continuous charge-density distribution over the primitive cell

$$\mathbf{P}_{\text{elec}} = \frac{1}{V_{\text{cell}}} \int_{\text{cell}} d\mathbf{r}\, \rho(\mathbf{r})\mathbf{r} \ , \tag{2.72}$$

where $\rho(\mathbf{r})$ is the real-space charge-density distribution of the valence electrons. However, in principle there are infinitely number of ways to choose a primitive cell, each of which would give us a unique result based on Eq. (2.72). Therefore, Eq. (2.72) is still problematic. There are also other attempts towards defining the electronic part of the polarization in terms of local charge-density distribution, as nicely summarized in Ref. [54], but all of them suffer from various fallacies.

On the other hand, from both experimental and theoretical points of view, it has been realized that the "absolute polarization" is ill defined [54, 63]. Experimentalists always measure the change of polarization in a ferroelectric material through a hysteresis loop. From a theoretical point of view, the bulk property of a crystal should remain invariant against the translation of one valence electron by one unit cell, but it may change the polarization by a "polarization quantum", which is simply one electric charge in 1D lattice [64]. It is certainly allowed to repeat such an operation for an arbitrary number of times. As a result, there are infinite number of polarization values that differ from each other by the polarization quanta, but all of them actually correspond to the same bulk physical state. An instantaneous polarization would never be unambiguously determined in this regard. As pointed out by Resta [65, 63], only the change of the polarization through an adiabatic process is an experimentally measurable quantity. Therefore it is more appropriate to define the electric polarization from the charge current instead of the charge density [65, 54]

$$\frac{d\mathbf{P}(t)}{dt} = \langle \mathbf{J}(t) \rangle \ , \tag{2.73}$$

where $\langle \mathbf{J}(t) \rangle$ is the quantum-mechanical expectation value of the current-density operator. If the current is driven by some external parameter λ that varies adiabatically with time, the change of the polarization from the initial ($\lambda = 0$) state to the final ($\lambda = 1$) state is then

$$\Delta \mathbf{P} = \int_0^1 d\lambda \, \langle \mathbf{J}(\lambda(t)) \rangle \ . \tag{2.74}$$

$\Delta \mathbf{P}$ turns out to be the measurable quantity without ambiguity.

The above idea motivated the development of the "modern theory of electric polarization" [14, 15]. Eq. (2.73) implies that the polarization is closely related to the phase of electronic wavefunctions. In 1993, King-Smith and Vanderbilt showed that during an adiabatic process parameterized by λ, the instantaneous polarization can be expressed as the Berry phase of the (occupied) instantaneous Bloch functions [14],

$$P_j(\lambda) = \frac{e}{(2\pi)^3} \sum_{n=1}^{N_{\text{occ}}} \int_{\text{BZ}} d\mathbf{k} \, i \, \langle u_{n\mathbf{k}}(\lambda) | \partial_j u_{n\mathbf{k}}(\lambda) \rangle \ , \tag{2.75}$$

where $\partial_j \equiv \partial / \partial k_j$, $u_{n\mathbf{k}}(\lambda)$ is the periodic part of the nth occupied Bloch function at $\mathbf{k}$, and N_{occ} is the number of occupied bands. We assume the system is a 3D lattice.

For simplicity, let us consider a 3D orthorhombic lattice [4]. It is helpful to write $\mathbf{k} = (\mathbf{k}_\parallel, k_j)$, where k_j represents the wavevector component in the direction of interest, and $\mathbf{k}_\parallel$ are the other two wavevector components which are orthogonal to k_j. Using these notations, Eq. (2.75) can be re-expressed as

$$P_j(\lambda) = \frac{e}{(2\pi)^3} \sum_{n=1}^{N_{\text{occ}}} \int_{\text{BZ}} d\mathbf{k}_\parallel \, \phi_{nj}(\mathbf{k}_\parallel, \lambda) \ , \tag{2.76}$$

where

$$\phi_{nj}(\mathbf{k}_\parallel, \lambda) = \int_0^{\frac{2\pi}{a_j}} dk_j \, i \, \langle u_{n\mathbf{k}}(\lambda) | \partial_j u_{n\mathbf{k}}(\lambda) \rangle \ , \tag{2.77}$$

is the Berry phase of $|u_{n\mathbf{k}}\rangle$ along the k_j direction at $\mathbf{k}_\parallel$. The total polarization is proportional to the sum of all the Berry phases at all $\mathbf{k}_\parallel$ in the 2D plane.

---

[4] Otherwise we must do the calculus in a skew coordinate system, which is more complicated. However, the general conclusion remains unchanged

The Berry-phase formula may also explain the ambiguity of the bulk polarization. It is rooted in the 2π ambiguity of the Berry phase. It follows from Eq. (2.76) that the magnitude of the polarization quantum in the jth direction is $e/A_0 = e\,a_j/V_{\rm cell}$, where $A_0 = V_{\rm cell}/a_j$ is the cross-sectional area of the primitive cell normal to $\mathbf{a}_j$. For crystals with non-orthogonal lattice vectors, the polarization quantum $\mathbf{P}_j^0$ in the direction of the jth lattice vector can be generalized as

$$\mathbf{P}_j^0 = e\,\frac{\mathbf{a}_j}{V_{\rm cell}}\ , \tag{2.78}$$

where $\mathbf{a}_j$ is the jth primitive lattice vector.

The generalization of Eq. (2.76) to 2D and 1D lattices is straightforward. For 2D systems,

$$P_j(\lambda) = \frac{e}{(2\pi)^2}\sum_{n=1}^{N_{\rm occ}}\int_{\rm BZ} dk_\parallel\,\phi_{nj}(k_\parallel,\lambda)\ , \tag{2.79}$$

where $k_\parallel$ becomes a scalar for 2D systems, and $\phi_{nj}(k_\parallel,\lambda)$ is just the Berry phase along the k_j direction as defined in Eq. (2.77). The polarization quantum in 2D is a generalization of Eq. (2.78),

$$\mathbf{P}_j^0 = e\,\frac{\mathbf{a}_j}{A_{\rm cell}}\ , \tag{2.80}$$

where $\mathbf{a}_j$ is one of the two lattice vectors, and $A_{\rm cell}$ is the area of the 2D primitive cell.

For 1D systems, the polarization has the same dimension as the electric charge,

$$P_j(\lambda) = \frac{e}{2\pi}\sum_{n=1}^{N_{\rm occ}}\phi_{nj}(\lambda)\ . \tag{2.81}$$

As mentioned above, the polarization quantum in 1D is simply the electric charge e.

The ambiguity of bulk polarizations is closely related to surface charges. Let us truncate a bulk 3D lattice to make a surface with the surface normal $\hat{\mathbf{n}}$. It follows from Eq. (2.78) that $\hat{\mathbf{n}}\cdot\mathbf{P}_j^0 = e/A_s$, where A_s is the area of the 2D surface primitive cell. It implies that the polarization quantum may be rooted in the uncertainty of the surface charge density. To be explicit, one can always add or remove an integer number of electrons from each surface primitive cell without affecting the bulk property, but

the bulk polarization would be changed by the corresponding polarization quanta due to such operations [15, 54]. Therefore, for an isolated insulating surface with definite surface charge density, the polarization would be uniquely determined by the surface condition.

The electric polarization can also be expressed in the Wannier-function basis. Using Eq. (2.31), the electric polarization for 3D systems can be re-expressed as

$$\mathbf{P} = \frac{e}{V_{\text{cell}}} \sum_{n=1}^{N_{\text{occ}}} \mathbf{r}_{n\mathbf{R}} \ , \tag{2.82}$$

where $\mathbf{r}_{n\mathbf{R}} = \langle w_{n\mathbf{R}} | \mathbf{r} | w_{n\mathbf{R}} \rangle = \langle w_{n\mathbf{0}} | \mathbf{r} | w_{n\mathbf{0}} \rangle + \mathbf{R}$ is the Wannier center (dependence on λ is implicit). Certainly the choice of the lattice vector $\mathbf{R}$ is arbitrary, corresponding to the ambiguity of the bulk polarization.

2.2.3 Anomalous Hall effect

Besides the electric polarization, another important application of the Berry phase in solids is the anomalous Hall effect. The linear response of electrons in solids to a weak applied electric field can be described by a conductivity (conductance) tensor for 3D (2D) systems

$$J_a = \sigma_{ab} E_b \tag{2.83}$$

where σ_{ab} refers to the conductivity (conductance) tensor, E_b is the electric field applied in the b direction, and J_a is the electric current density in the a direction induced by E_b. If $a = b$, the above formula is basically Ohm's law in metals. If $a \neq b$, Eq. (2.83) describes a dissipationless transverse current generated by a longitudinal electric field in the absence of external magnetic fields. Such an effect is known as the anomalous Hall effect (AHE), and σ_{ab} is known as the anomalous Hall conductivity (conductance). In 2D or quasi-2D systems, the anomalous Hall conductance is usually denoted as $\sigma_{xy} = -\sigma_{yx}$. In 3D systems, there are three anomalous Hall conductivities denoted by σ_{xy}, σ_{zx} and σ_{yz}.

It turns out that there are three main mechanisms which add up to contribute to the total AHE, as nicely summarized in Ref. [16]. First, in realistic materials with impurities and defects, there would be skew scattering of the spin-polarized electrons by disorder, which results from the spin-orbit coupling of either the intrinsic system or the disorder. Second, there is also the "side-jump" mechanism which takes account of the other effects from imperfections of the lattice. Lastly, besides these "extrinsic" mechanisms, there is an intrinsic contribution which tends to dominate in disordered systems at relatively high temperatures.

The intrinsic contribution to the AHE was first pointed out about 60 years ago by Karplus and Luttinger [66] in terms of the "anomalous velocity". A half century later, after the idea of Berry's geometric phase was proposed [11], the anomalous-velocity term was reformulated as the Berry curvature of the semiclassical wavepacket subject to a weak electric field [67]. In 2002, Jungwirth *et al.* [68] studied the AHE in dilute ferromagnetic semiconductors, where the anomalous Hall conductivity (conductance) was expressed as the integration of the Berry curvature of Bloch bands over the occupied portion of the BZ

$$\sigma_{ab} = \frac{e^2}{\hbar} \sum_n \int_{\mathrm{BZ}} \frac{d\mathbf{k}^d}{(2\pi)^d} \Omega_{ab,n}(\mathbf{k}) f_n(\mathbf{k}, \mu) \; . \tag{2.84}$$

In the above equation, d refers to the spatial dimension, $\Omega_{ab,n}$ is the Berry curvature of the nth Bloch band, and $f_n(\mathbf{k}, \mu)$ is the occupation number of the nth band at $\mathbf{k}$ with the Fermi level μ.

If the system has $\mathcal{T}$ symmetry, from Sec. 2.1.4 we know that the traces of Berry curvatures at opposite $\mathbf{k}$ points are opposite, i.e., $\sum_n \Omega_{ab,n}(\mathbf{k}) = -\sum_n \Omega_{ab,n}(-\mathbf{k})$ (Eq. (2.41) and Eq. (2.50)). On the contrary, the occupation number at $\mathbf{k}$ and $-\mathbf{k}$ are equal to each other: $f_n(\mathbf{k}) = f_n(-\mathbf{k})$. Therefore, σ_{ab} must vanish for a system with $\mathcal{T}$ symmetry. The breaking of $\mathcal{T}$ symmetry thus becomes necessary to obtain the AHE. On the other hand, in typical magnetic materials, the magnetic moments are mostly contributed by the spin degrees of freedom, which means $\mathcal{T}$ symmetry is broken

in the electronic spin subspace. The Berry curvature is associated with orbital motion of electrons, which implies the spin-orbit coupling is also indispensable for the AHE.

For 2D metals, Haldane showed that Eq. (2.84) can be expressed as the Berry phase of the Bloch states around the Fermi loop(s) [17],

$$\begin{aligned}
\sigma_{xy} &= \frac{e^2}{h}\frac{1}{2\pi}\sum_n \int_{\mathrm{BZ}} d\mathbf{k}^2 \Omega_{n,xy}(\mathbf{k}) f_n(\mathbf{k},\mu) \\
&= \frac{e^2}{h}\frac{1}{2\pi}\sum_n \int_{\mathrm{BZ}} d\mathbf{k}^2 \Big(A_{n,x}(\mathbf{k})\partial_y f_n(\mathbf{k},\mu) - A_{n,y}(\mathbf{k})\partial_x f_n(\mathbf{k},\mu) \Big) \\
&= \frac{e^2}{h}\frac{1}{2\pi}\oint_{\mathcal{C}_F} d\mathbf{k}_F \cdot \mathbf{A}(\mathbf{k}_F) \\
&= \frac{e^2}{h}\frac{\beta(\mathcal{C}_F)}{2\pi} \,, \qquad (2.85)
\end{aligned}$$

where $\mathcal{C}_F$ refers to the Fermi loop(s) of the 2D metallic system, $\mathbf{k}_F$ is the Fermi wavevector, $\mathbf{A}(\mathbf{k}_F)$ is the corresponding Berry connection at $\mathbf{k}_F$, and $\beta(\mathcal{C}_F)$ denotes the Berry phase around the Fermi loop(s). The second line of Eq. (2.85) follows by using the integration by parts. The third line follows from the fact that at zero temperature the partial derivative of $f_n(\mathbf{k},\mu)$ with respect to $\mathbf{k}$ behaves as a δ-function exactly at the Fermi wavevector $\mathbf{k}_F$, and vanishes identically elsewhere.

For 3D metallic systems, Eq. (2.85) may be generalized as a Fermi-surface integral of the Berry curvature weighted by the corresponding local Fermi wavevector, plus an additional term from the BZ boundary [17]. However, for the purpose of practical calculations, it may be more intuitive to follow the idea of Ref. [69], in which the 3D BZ is divided into a discrete set of 2D parallel $\mathbf{k}$ slices, and the anomalous Hall conductivity is proportional to the sum of the Berry phases around all the Fermi loops on all the slices. To be specific, following the notation of Ref. [69], the anomalous Hall conductivity of 3D metals can be expressed as

$$\sigma_{ab} = \frac{e^2}{h}\frac{1}{2\pi}\sum_{n,c} \epsilon_{abc} K_{n,c} \,, \qquad (2.86)$$

where

$$\begin{aligned}
\mathbf{K}_n &= \sum_{i=1}^{3} g_{ni}\mathbf{G}_i \\
g_{ni} &= \frac{a_i}{(2\pi)^2}\int_0^{2\pi/a_i} dk_i \beta_n(k_i) \\
\beta_n(k_i) &= \oint_{\mathcal{C}_n(k_i)} d\mathbf{k}\cdot\mathbf{A}(\mathbf{k}) \ .
\end{aligned} \tag{2.87}$$

In the above equation, $\mathbf{G}_i$ denotes the ith reciprocal lattice vector, a_i is the lattice constant in the ith direction, $\mathcal{C}_n(k_i)$ is the projected Fermi loop(s) within the 2D slice at k_i for the nth band (k_i is the direction normal to the 2D $\mathbf{k}$ slice), and $\beta_n(k_i)$ is the Berry phase of the nth band around $\mathcal{C}_n(k_i)$.

If band n is completely filled, then there is no Fermi surface nor any projected Fermi loop. The Berry phase $\phi_n(k_i)$ thus must be evaluated over the entire 2D $\mathbf{k}$ slice, which equals to integer multiples of 2π:

$$\begin{aligned}
\beta_n(k_i) &= 2\pi C_n^i \ , \\
g_{ni} &= C_n^i \ ,
\end{aligned} \tag{2.88}$$

where the integer C_n^i is the Chern number of the nth band defined in the 2D slice at k_i. Therefore, the contribution to the total anomalous Hall conductivity from a completely filled band must be quantized in terms of the reciprocal lattice vectors,

$$\sigma_{ab}^n = \frac{e^2}{h}\frac{1}{2\pi}\sum_c \epsilon_{abc} G_{n,c} \ , \tag{2.89}$$

where $\mathbf{G}_n = \sum_{i=1}^3 C_n^i \mathbf{G}_i$. It follows that all the completely filled bands would contribute to the total σ_{ab} as a quantized part, while the remaining non-quantized part is a pure Fermi-surface property as shown in Eq. (2.87).

As a concluding remark, we would like to briefly discuss the situation of insulators. From Eq. (2.85) and Eq. (2.87), we know that the non-quantized part of the anomalous Hall conductivity (conductance) can be expressed as a Fermi-surface (Fermi-loop) property in 3D (2D) systems, and that a completely filled band may only contribute

a quantized anomalous Hall response. If the system is insulating (say, in two dimensions), the non-quantized part must vanish because there is no Fermi loop. Therefore the anomalous Hall conductance must be quantized in units of e^2/h, $\sigma_{ab} = Ce^2/h$. The integer C is the Chern number of the occupied bands for the 2D insulating system. We call such a 2D insulator with non-zero Chern number a quantum anomalous Hall insulator, or Chern insulator [24]. We will discuss the quantum anomalous Hall insulator in more detail in Sec. 2.3.2.

2.2.4 Chern-Simons orbital magnetoelectric coupling

Another important property of solids which involves the geometric phase of the Bloch functions is the magnetoelectric coupling. With magnetoelectric (ME) coupling in an insulator, an electric polarization $\mathbf{P}$ can be induced by an external magnetic field $\mathbf{B}$, and conversely a magnetization $\mathbf{M}$ can be generated by an applied electric field $\mathbf{E}$. The linear ME coupling coefficient is a rank-2 tensor defined as

$$\alpha_{ab} = \frac{\partial M_b}{\partial E_a}\Big|_{\mathbf{E}=0} = \frac{\partial P_a}{\partial B_b}\Big|_{\mathbf{B}=0} \tag{2.90}$$

where a, $b = x, y, z$ denote the directions in real space. The ME phenomena are contributed by both electronic and lattice degrees of freedom. To be specific, the electronic contribution refers to the ME response when the ions are completely frozen, while the lattice contribution takes account of the ionic response of $\mathbf{M}$ ($\mathbf{P}$) to external electric (magnetic) fields. Moreover, depending on the origin of the $\mathbf{E}$-induced magnetization, each of the two contributions can be further decomposed into spin and orbital components [70, 71].

The spin contribution to the ME response (from both electronic and lattice degrees of freedom) has been thoroughly studied with well established theoretical methods in typical magnetoelectrics such as Cr_2O_3 [72, 73]. On the other hand, the orbital ME response is theoretically more challenging and intriguing. In recent years, it has been shown that the frozen-ion orbital ME coupling consists of two terms. One term can

be expressed as a standard linear response of the Bloch functions to external electric (magnetic) fields, denoted as the "Kubo term"; while the other is isotropic and is completely determined by the unperturbed ground-state wavefunctions, which is known as the Chern-Simons term [71, 74].

The Chern-Simons coupling in insulators is conventionally scaled by a dimensionless parameter θ

$$\alpha_{ij}^{\mathrm{CS}} = \frac{\theta e^2}{2\pi h}\delta_{ij} \ , \tag{2.91}$$

where θ is another type of geometric phase of the occupied Bloch states, which can be expressed as an integral of the Chern-Simons 3-form over the 3D BZ [19, 20]

$$\theta = -\frac{1}{4\pi}\int d^3k \ \epsilon_{abc}\mathrm{Tr}\left[A_a\partial_b A_c - \frac{2}{3}iA_aA_bA_c\right] , \tag{2.92}$$

where A_a, A_b and A_c are the Berry connections of the occupied Bloch states, and the trace is taken over the occupied bands. The Berry phase defined in Eq. (2.27) can be considered as the integral of the Berry connection, which is the Chern-Simons 1-form, over the 1D BZ. θ may be interpreted as the generalization of the Berry phase to 3D manifolds, i.e., it is the integral of the Chern-Simons 3-form over the 3D BZ.

The Chern-Simons ME coupling has several interesting properties. First of all, a material with the Chern-Simons ME coupling can be considered as medium exhibiting axion electrodynamics [75], where an additional term $\Delta\mathcal{L} = \alpha^{\mathrm{CS}}\mathbf{E}\cdot\mathbf{B}$ is added to the conventional Lagrangian of electromagnetic fields in media. The electrodynamics with such an axion coupling turns out to be invariant under $\theta \to \theta + 2\pi$ [75].

Secondly, θ would have a physically measurable effect only if it varies in space or time [74]. As a result, for a time-independent crystal with a surface truncation, the presence of the bulk Chern-Simons coupling is manifested as the surface anomalous Hall effect, where the anomalous Hall conductance is proportional to θ, $\sigma_{xy} = \theta e^2/(2\pi h)$. The connection between the surface anomalous Hall effect and the bulk Chern-Simons ME coupling may provide an intuitive explanation to the ambiguity of θ. If an insulating

quantum anomalous Hall (QAH) layer with non-zero Chern number C is coated to all the surfaces of a 3D system with non-zero θ response in the bulk, the surface anomalous Hall conductance would be changed by the Chern number $\sigma_{xy} = \theta e^2/(2\pi h) + Ce^2/h$, while the bulk property remains unchanged. Such a freedom of coating QAH layers with arbitrary Chern numbers to the surface would take account for the 2π ambiguity in defining θ. It is interesting to note that the ambiguity of θ is analogous to the ambiguity in defining the bulk electric polarization as discussed in the previous subsection. In the latter case, the ambiguity is due to the freedom of adding (removing) an integer number of charges to (from) the surface bands [15].

The 2π ambiguity of θ imposes a $\mathbb{Z}_2$ topological classification to 3D insulators with either time-reversal or inversion symmetry. Both $\mathcal{T}$ and $\mathcal{P}$ operations would transform θ to $-\theta$, while the presence of $\mathcal{T}$ or $\mathcal{P}$ symmetry implies that $\theta = -\theta$, modulo 2π. Therefore, there are only two possible values of θ, either 0 or π. When $\theta = \pi$, the system is known as a topological insulator [33, 34] in the presence of $\mathcal{T}$ symmetry; while it is known as an axion insulator when only inversion symmetry is present [76]. We will discuss about some of the exotic topological phases in the following section.

2.3 From geometry to topology

In Sec. 2.1, we have introduced the concept of Berry phase and discussed various properties of the Berry phase. We also showed a few examples to illustrate how the geometric phases are involved in specific physical problems. We see that the electronic part of the polarization in insulating solids can be expressed as some kind of averaged Berry phase of the occupied Bloch states; the anomalous Hall conductances in (quasi) 2D metallic systems are proportional to the Berry phases of the Bloch states around the Fermi loops; for 3D lattices, the Chern-Simons orbital ME coupling is proportional to another geometric phase θ, which is the integral of the Chern-Simons 3-form over the entire 3D BZ.

These examples indicate that some physical response functions may be determined by the geometric properties of the ground-state wavefunctions. The values of these geometric quantities are strongly dependent on the specific properties of the systems, which are not necessarily quantized. For example, in a typical ferroelectric material such as $BaTiO_3$, the bulk polarization would take some definite value modulo the polarization quantum, but in general it differs from that of another type of ferroelectrics such as $PbTiO_3$. The anomalous Hall conductivity of Fe is definitely different from that of Co, even though they are next to each other in the periodic table.

In some special situations, however, such geometric-phase-related quantities may be exactly quantized regardless of specific properties of the materials. For examples, it is possible to have a (quasi) 2D system whose anomalous Hall conductance is exactly quantized in units of e^2/h, which is robust against weak perturbations. In topological insulators (axion insulators), the formal bulk ME responses are always quantized as $e^2/2h$, as long as time-reversal (inversion) symmetry is preserved. In such cases, the exact quantization of the physical response functions are protected by the corresponding topological invariants. Therefore they are insensitive to the details of the systems.

In this section, we will discuss some of the topologically protected phases. We start with the first discovered topological phases of matter, the integer quantum Hall effect (IQH) [21], and extend the discussion to the quantized Hall effect without external magnetic field, i.e., the quantum anomalous Hall (QAH) effect [24, 25]. We proceed to the situation with preserved $\mathcal{T}$ symmetry, and introduce the quantum spin Hall (QSH) effect [27, 26, 28] and the topological insulators [33, 34].

2.3.1 The integer quantum Hall effect

In 1980, von Klitzing discovered that the Hall conductance of a 2D electron gas subject to strong out-of-plane magnetic fields was exactly quantized in units of e^2/h at low temperatures, $\sigma_{xy} = \nu e^2/h$ [21]. This is known as the IQH effect.

It is not difficult to explain the IQH effect if the system is modeled as a clean free-electron gas subject to strong magnetic fields, where the effects of ionic potentials, defects, impurities and electron-electron interactions are all ignored. To be specific, let us consider a model Hamiltonian of a 2D free electron gas with a strong applied magnetic field of magnitude B in the z direction and a weak applied electric field of magnitude E in the longitudinal direction (y direction). If we take the gauge choice $\mathbf{A} = (0, Bx - Et, 0)$, and ignore the Zeeman splitting of the spins [5], the Hamiltonian looks like

$$H = \frac{1}{2m_e}\Big((-i\hbar\frac{\partial}{\partial_x})^2 + (-i\hbar\frac{\partial}{\partial_y} + eBx - eEt)^2\Big) \ . \tag{2.93}$$

Suppose the electric field is infinitesimally weak, so that we can ignore its influence on the energy spectrum. If the electric field is temporarily dropped, the eigenenergies of the Hamiltonian shown in Eq. (2.93) form Landau levels $E_n = (n + 1/2)\,\hbar\omega_c$, where $\omega_c = eB/m_e$ (with m_e being the mass of an electron). Each Landau level turns out to be highly degenerate. The degeneracy $D = SeB/h$, where S is the area of the entire 2D sample. The corresponding eigenfunction $\psi_n(x, y)$ is plane-wave-like in the y direction, and can be written in the form $\psi_n(x, y) = e^{ik_y y}\phi_n(x - x_c)$, where $\phi_n(x - x_c)$ is some wavepacket centered at $x_c = -\hbar k_y/eB$. The effect of the electric field may be taken into account by the Pierel's substitution $k_y \to k_y - eEt/\hbar$. Then the center of the wavepacket becomes time-dependent $x_c(t) = -\hbar k_y/eB + Et/B$, with the group velocity $v_c = dx_c/dt = E/B$. When the chemical potential lies in the gap between two Landau levels, the current density along x direction $j_x = -e\rho v_c$, where $\rho = \nu D/S = \nu eB/h$ is the number of electrons per unit area. Combining the above two equations, one immediate obtains $j_x = -\nu e^2 E/h$, and $\sigma_{xy} = -\nu e^2/h$, where ν is the number of occupied Landau levels.

---

[5] For a magnetic field as strong as 10T, the corresponding Zeeman splitting is $\sim$ 1meV, which is negligible compared with the Landau-level splitting. So the Zeeman splitting will be ignored in the following discussions.

From the above discussions, it seems that the IQH effect naturally results if an integer number of Landau levels are filled. However, the above analysis is based on an over-simplified model Hamiltonian as shown in Eq. (2.93). In realistic systems, there are Coulomb interactions between electrons (which should be weak for IQH effct), nuclear potentials, and lattice imperfections such as impurities and defects. The realistic Hamiltonian describing the experimental system is far more complicated, and is essentially disordered. Then how the exact quantization of the Hall conductance remains robust against these perturbations?

There are a lot of theoretical works aiming to prove the quantization of the Hall conductance in the presence of various types of disorder, as briefly summarized in Chapter 13 of Ref. [61]. One of the most important conclusions from these works is that each Landau level would be broadened into Landau bands with finite bandwidth in the presence of disorder. Moreover, the states around the center of the Landau level remain extended in the presence of disorder, while they become localized as the energy approaches the band edges. The extended states and localized states are separated from each other by a "mobility edge", whose exact position is dependent on the strength and types of disorder, as well as specific properties of the system. Based on the above conclusion, Thouless managed to prove that the Hall conductance would remain exactly quantized at zero temperature as long as the Fermi level lies in the "mobility gap" [77] [6]. Interestingly, by modelling the potential of a single impurity as a δ-function in real space, Prange proved that even though the localized states around the single impurity carries no current, there is an extra Hall current from the extended states which exactly compensate for the loss of current from the localized state. Therefore the exact quantization of the Hall conductance remains robust [78].

Motivated by the previous studies, Laughlin proposed his famous gauge argument

[6] Here the mobility gap means an energy interval within which there are only localized eigenstates, which do not contribute to the transport phenomena.

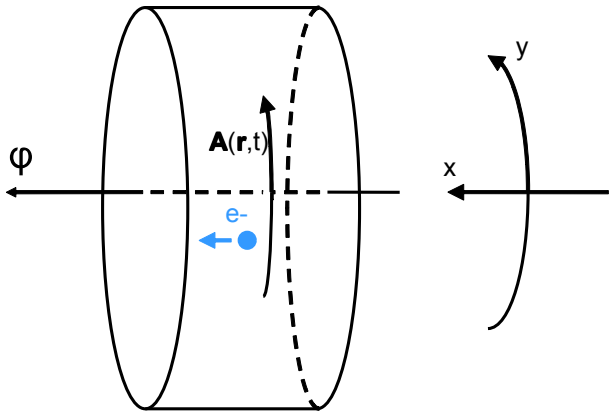

Figure 2.2: A schematic plot of the Laughlin's ribbon. A gauge field $\mathbf{A}(\mathbf{r}, t)$, which is dependent on both position $\mathbf{r}$ and time t, is applied toward the longitudinal direction. As a result of the longitudinal gauge field, a magnetic flux ϕ is threaded through the ribbon, and a constant magnetic perpendicular to the surface of the ribbon is applied.

[79]. He considered a 2D Hall-bar system with the periodic boundary condition in the y direction (longitudinal) and the open boundary condition in the x direction (transverse). Thus a Hall bar is modelled as a ribbon of finite width as shown in Fig. 2.2. A uniform magnetic field perpendicular to the surface of the ribbon is applied, and a magnetic flux ϕ that varies slowly with time threads through the ribbon as shown in Fig. 2.2. It is convenient to chose the gauge $\mathbf{A} = (0, Bx + \phi_0 t/(\tau L), 0)$, where B is the magnitude of the magnetic field, $\phi_0 = h/e$ is the flux quantum, τ is the period of time after which ϕ varies by a flux quantum, and L is the circumference of the ribbon. We see that A_y consists of two parts [7]. The spatially dependent part generates a constant magnetic field normal to the surface of the ribbon, while the time-dependent part threads a time-dependent flux through the ribbon. As time varies from 0 to τ (consider $\tau \to \infty$), the flux ϕ increases linearly from 0 to ϕ_0.

As the flux increases adiabatically, a longitudinal current would be generated [80]

$$S\, J_y = \frac{e}{m_e} \sum_{j=1}^{N} (-i\hbar\, \partial_{y_j} - eA_y(\mathbf{r}_j)) = -\frac{\partial H(t)}{\partial A_0(t)}\Big|_{A_0(t)=0}, \tag{2.94}$$

where J_y is the current density operator in the longitudinal direction, $A_0(t) = \phi_0 t/(\tau L)$,

[7] In the rectangular geometry, the longitudinal direction is the y direction. But in the ribbon geometry, it is the azimuthal direction. We just label the azimuthal direction as the "y" direction for convenience.

$S = L\,W$ is the area of the sample (W is the width of the ribbon in the transverse direction and L is the circumference of the ribbon), j is the index of electrons, and the total number of electrons is N (a macroscopically large number). From Eq. (2.94), it follows that the expectation value of the longitudinal current $I_y = \langle\, J_y \,\rangle\, W$ can be expressed as

$$\begin{aligned} I_y &= -\langle\, \frac{\partial H(\phi(t))}{\partial \phi(t)} \,\rangle\Big|_{\phi(t)=0} \\ &= -\frac{\partial E(\phi(t))}{\partial \phi(t)}\Big|_{\phi(t)=0} \\ &= -\frac{\Delta E}{\Delta \phi} \end{aligned} \tag{2.95}$$

where $\phi(t) = A_0(t)L = \phi_0 t/\tau$. The second line in the above equation follows from the Hellmann-Feynman theorem, and in the last line the continuous derivative is approximated by a finite difference.

Consider the process that ϕ increases from 0 to ϕ_0, i.e., $\Delta\phi = \phi_0 = h/e$. After such a process, the energy must increase by some amount to generate a non-vanishing longitudinal current. However, the eigenstates at $\phi = \phi_0$ must remain the same as those at $\phi = 0$. The only way that the energy could increase is "through a re-population of the states" [80]. On the other hand, if there is no extended bulk states at the Fermi level, in other words, the Fermi level lies in the (mobility gap) gap between two Landau levels (bands), the population of electrons in the bulk must remain invariant after such an adiabatic cycle. Then the only possible way to increase the energy is to pump an exactly integer number (ν) of electrons from one edge to the other. In other words, the energy increases through a re-population of the edge states. Therefore, $\Delta E = \nu e \Delta V$ (ν is an integer), where ΔV is the potential drop across the transverse direction of the ribbon. It immediately follows that $I_y = -\Delta E/\Delta\phi = -\nu e^2/h\,\Delta V$, i.e., $\sigma_{yx} = -\sigma_{xy} = -\nu e^2/h$. Laughlin further showed that in the clean limit, ν is exactly the number of the filled Landau levels [80].

We would like to make two additional remarks about Laughlin's argument. First, it

should be noted that Laughlin's gauge argument is model-independent. The only two requirements to validate the argument is the correctness of Eq. (2.94), and that there is no bulk extended states at the Fermi level. Therefore, it is fair to claim that the quantization of the Hall conductance remains exact against various perturbations, as along as the perturbations are not strong enough to close the bulk energy gap (either a spectrum gap or a mobility gap). From this point of view, it seems that the quantized Hall conductance is topologically protected by a finite energy (mobility) gap.

Secondly, the adiabatic charge pump from one edge to the other suggest that there must be some gapless extended states localized at the two edges. Following Laughlin's work, Halperin in 1982 obtained the exact solution of a 2D free-electron system in the annular geometry subject to a strong out-of-plane magnetic field [81]. It was shown that the eigenenergies remain gapped and approach the limit of flat Landau levels in the bulk, while they increase monotonically and become gapless as one approaches the edges. It was further demonstrated that these gapless "chiral" edge states carry opposite currents. All the states are left-moving at one edge, while they become all right-moving at the opposite edge. In principle, extended electronic states in 1D systems (the edges of a 2D systems are considered as 1D systems) would be localized by arbitrarily weak disorder due to the backscattering off the disordered potentials. For such chiral edge states, however, there is no state available for the backscattering. Therefore these edge states remain robust even in the presence of moderate strength of disorder, which were argued to be responsible for the quantized Hall conductance [81]. This is the first time that the concept of topologically protected edge states was explicitly proposed in condensed matter physics.

The topological nature of the IQH effect was later on clarified by Thouless, Kohmoto, Nightingale, and den Nijs [22]. Thouless *et al.* explicitly calculated the bulk Hall conductance for 2D non-interacting electrons subject to a periodic potential and a perpendicular magnetic field. When the strength of the magnetic field is chosen in such

a way that an integer number (p) of magnetic flux quanta penetrate through an integer number (q) of primitive cells (in general $p \neq q$), it turns out that the eigenfunctions of the Hamiltonian are still Bloch-like. A "magnetic BZ", whose size is q-times smaller than the BZ without magnetic field, can be defined [23]. For example, for a gauge choice $\mathbf{A} = (0, Bx, 0)$, the wavevector in the magnetic BZ may be defined as $k_x \in (0, 2\pi/(qa)]$, $k_y \in (0, 2\pi/b]$, where a and b are the real-space lattice constants of the original crystal in the x and y directions respectively [22].

Using the Kubo formula, it turns out that when the Fermi level lies in a gap, the Hall conductance in such a probelm can be written as the integral of the Berry curvature of the Bloch-like functions over the entire magnetic BZ (MBZ) [22, 23],

$$\sigma_{xy} = \frac{e^2}{2\pi h} \int_{\mathrm{MBZ}} dk_x dk_y \sum_n \Omega_{xy,n}(\mathbf{k}) \ , \tag{2.96}$$

where $\Omega_{xy,n}(\mathbf{k})$ denotes the Berry curvature defined in terms of the periodic part of the quasi-Bloch functions. $\mathbf{k} = (k_x, k_y)$ are defined in the MBZ, while n labels the occupied bands. As mentioned in Sec. 2.2.3, the integral of the Berry curvature over a closed 2D parameter space must be quantized as $2\pi C$, where C is known as the Chern number. In the context of the IQH effect, the parameter space is the 2D magnetic BZ, and the Berry curvature is calculated with respect to the occupied Landau bands. Therefore, the quantization number in IQH effect $\nu = \sum_n C_n$, i.e., the sum of the Chern numbers of all the occupied bands. The Chern number defined in the magnetic BZ is also known as the TKNN invariant in the literatures.

The topological nature of the IQH effect naturally explains the robustness of the quantization against various perturbations. As shown in Eq. (2.96), the Hall conductance is directly proportional to the TKNN invariant, while the TKNN invariant is guaranteed to be quantized as long as the bulk energy gap (which could be a mobility gap) remains unclosed. Therefore, any weak perturbations, e.g., electron-electron interactions, impurities, defects, which are not strong enough to close the bulk energy gap would not change the quantized Hall conductance. Moreover, as shown by Halperin

[81], a non-zero bulk Chern number (TKNN invariant), is always accompanied by the same number of gapless topological edge states. As discussed above, these edge states are chiral, in a sense that the group velocities of the gapless states localized at one edge all point to the same direction, thus eliminating the possibility of backscattering.

The most important concepts in (non-interacting) topological phases of matter are very well demonstrated in the IQH effect. First, we see that a bulk topological invariant can be defined in a gapped system. A bulk topological index is further associated with some bulk response function, leading to the exact quantization of the corresponding response function. In this regard, we say that the quantized response function is protected by the topologically nontrivial bulk energy gap. In the context of the IQH effect, the bulk energy gap is the gap between adjacent Landau levels or Landau bands, and the topological index is the Chern number, or the TKNN invariant, which is defined in terms of the occupied quasi-Bloch functions over the magnetic BZ. The response function associated with the topological index is the Hall conductance. More interestingly, a non-trivial bulk topological index in a d-dimensional system always corresponds to some gapless surface states [8] that cannot exist in any isolated $(d-1)$-dimensional systems [34]. For example, in the IQH systems, the topological surface states are the chiral gapless states localized at the two edges, which were never found in isolated 1D systems. In the following subsections, these concepts will be applied to a family of topological systems which are all rooted in the IQH effect.

2.3.2 The quantum anomalous Hall effect

After the discovery of the IQH effect, people started thinking about the possibility of achieving a spontaneous quantized Hall conductance in the absence of any external magnetic field. Such an effect is known as the quantum anomalous Hall (QAH) effect. The system exhibiting such a spontaneous quantized Hall effect is known as a QAH

[8] Here "surface" refers to the $(d-1)$-dimensional boundary of a d-dimensional system.

insulator, or a "Chern insulator".

In Sec. 2.2.3, we have discussed the anomalous Hall effect in metals with broken $\mathcal{T}$ symmetry. As shown in Eq. (2.85), the 2D Hall conductance in metallic systems is expressed as the Berry phases around the Fermi loops in the 2D BZ. On the other hand, in the previous subsection we also introduced the concept of bulk topological indices (the TKNN invariant, or the Chern number defined in the magnetic BZ) in gapped systems. The Chern insulators may thus be interpreted from two perspectives. First, it may be considered as a generalization of the AHE in metals in the sense that the integration of the Berry curvature has to be carried out over the entire BZ, because the system is an insulator instead of a metal. Second, it can be regarded as a generalization of the IQH effect in the absence of magnetic field. Therefore, the anomalous Hall conductance for a Chern insulator is quantized in terms of the Chern number of the occupied Bloch bands,

$$\sigma_{xy} = C\frac{e^2}{h} \tag{2.97}$$

where the Chern number C is defined as the integral of the Berry curvature over the entire BZ of the 2D lattice system

$$\begin{aligned} C &= \frac{1}{2\pi}\int_{\mathrm{BZ}} dk_x dk_y \,\mathrm{Tr}\left[\,\Omega_{xy}(\mathbf{k})\,\right] \\ &= \frac{1}{2\pi}\int_{\mathrm{BZ}} dk_x dk_y \sum_n \Omega_{xy,n}(\mathbf{k}) \end{aligned} \tag{2.98}$$

where the non-Abelian Berry curvature of the occupied Bloch bands Ω_{xy} is defined in Eq. (2.32), and n is the index of the occupied bands. Note the difference between Eq. (2.97-2.98) and Eq. (2.96). In the latter case, the integration is carried out over the magnetic BZ, and the Berry curvature is defined with respect to the quasi-Bloch functions which are eigenfunctions of "magnetic translation operators" [23, 61], instead of the regular Bloch functions in periodic lattices.

Similar to the IQH effect, the existence of chiral gapless edge states is also the benchmark of a Chern insulator [24, 82]. As discussed in the previous subsection,

these chiral edge states are free from backscattering, thus remain robust against weak disorder. The edge states on the two opposite edges carry opposite currents, both of which are dissipationless due to their chiral nature.

The Chern insulator was first theoretically proposed by Haldane in 1988 [24]. Haldane designed a tight-binding model of spinless electrons with broken $\mathcal{T}$ symmetry on a 2D honeycomb lattice. In this model, there is a real first-neighbor hopping term, a complex second-neighbor hopping, and staggered on-site energies which are opposite for the two sublattices,

$$H = t_1 \sum_{\langle i,j \rangle} (c_i^\dagger c_j + c.c.) + t_2 \sum_{\langle\langle i,j \rangle\rangle} e^{i\phi_{ij}} (c_i^\dagger c_j + c.c.) + \Delta \sum_i (-1)^\eta c_i^\dagger c_i \ . \tag{2.99}$$

As schematically shown in Fig. 2.3(a), t_1 and t_2 are the first-neighbor and second-neighbor hopping amplitudes, and Δ denotes the magnitude of the on-site energies. $\phi_{ij} = \pm\phi$ is a phase angle associated with t_2, whose sign is dependent on the chirality of the second-neighbor bond and cannot be made vanishing by a gauge transformation. The non-vanishing phase angle ϕ generates some microscopic magnetic flux whose average over a primitive cell vanishes.

It is straightforward to solve Eq. (2.99) and calculate the Berry curvature for the band lower in energy. Eq. (2.99) is written in the basis of two orbitals localized at A and B sublattices, therefore there are two energy bands in total. We consider the situation of hall filling, and calculate the Chern number of the lower (occupied) band as a function of the parameters of the model. Setting $t_1 = 1$ and $t_2 = 1/3$, Fig. 2.3(b) demonstrates the phase diagram of the Haldane model in the parameter space of Δ and ϕ. Given the value of Δ, say, $\Delta = 2$, if $\phi = 0$, the Chern number of the system must vanish as a result of $\mathcal{T}$ symmetry. As ϕ increases from 0, $\mathcal{T}$ symmetry is broken, and the bulk gap at one of the BZ corners (labeled by K) gradually diminishes. At some critical point ϕ_c, the gap at K closes and the system becomes a 2D Dirac semimetal. If $\phi > \phi_c$, the gap at K would be reopened in an "inverted" order, giving rise to a non-zero

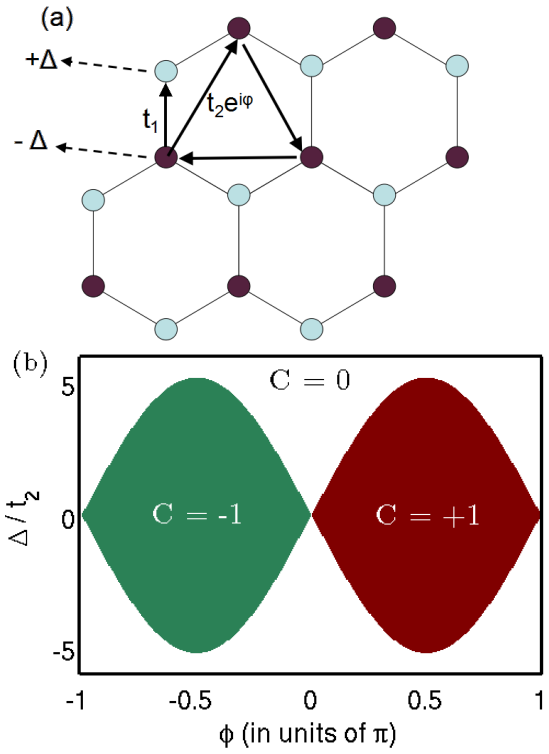

Figure 2.3: (a) The relevant parameters in Haldane model. (b) The phase diagram of the Haldane model.

Chern number. The system thus enters the Chern-insulator phase with a quantized Hall conductance proportional to the Chern number ($C = 1$). As ϕ keep increases, there is another gap closure at K so that the states return to the previous normal order and the Chern number vanishes again.

It is helpful to understand the topological character of the Chern insulator by looking at the "hybrid Wannier centers". To be explicit, consider the 2D system as a family of 1D chains parameterized by k_y, then we ask how the centers of the Wannier functions localized in the x direction (Eq. (2.31)) evolve as a function of k_y ($x_n(k_y)$). Such a quantity is called hybrid Wannier center because it can be computed by Fourier transforming a 2D Bloch function only in the x direction, such that the obtained function is Wannier-like in the x direction but remains Bloch-like in the y direction.

In the same vein as Laughlin's gauge argument, let us consider the situation that the quasi-1D chains grown along the x direction are wrapped around in the y direction,

forming "quasi-1D tubes". A magnetic flux is then threaded through the atomically thin quasi-1D tube, corresponding to a gauge field pointing to the y (or azimuthal) direction. The gauge field is spatially uniform [9] but linearly dependent on time. According to the Pierel's substitution, a change of the gauge field may be regarded as a change of the wavevector in the longitudinal direction (k_y). Therefore, the change of the flux by ϕ_0 (flux quantum) may correspond to the change of k_y by 2π, both of which leave the bulk of the quasi-1D system invariant. However, during such adiabatic evolution, an integer number of charges in the bulk of the quasi-1D system may be pumped by one lattice constant in the transverse (x) direction to the ends of the system, leading to excitations at the two ends.

Such a topological charge pump is well depicted by the hybrid Wannier centers. For a normal 2D insulator with vanishing Chern number, $x_n(k_y)$ would return to the original position as k_y varies by 2π, as shown by the black curve in Fig. 2.4. On the other hand, for a Chern insulator with the Chern number $C = 1(-1)$, $x_n(k_y)$ would be shifted by exactly one lattice constant along the positive (negative) x direction, as illustrated by the red crosses in Fig. 2.4. The data shown in Fig. 2.4 are obtained from the Haldane model in the topologically trivial and nontrivial phases respectively.

As a result of the non-trivial bulk topology, electrons from the bulk of the quasi-1D system are pumped to the state localized at the end whose energy level increases with k_y. On the other hand, the lowest unoccupied state localized at the opposite end of the chain is pulled down due to the charge pump. After one adiabatic cycle (k_y changes by 2π), an exactly integer number of electrons are accumulated at one end with the same number of holes left at the other end. If one plots the dispersion of the energy level localized at one end with respect to k_y, one would obtain some chiral states traversing through the bulk energy gap as shown in Fig. 2.5. These states are exactly the gapless

[9] In the case of QAH effect, an external magnetic field is not needed. Therefore, we can drop the spatial dependence of the gauge field.

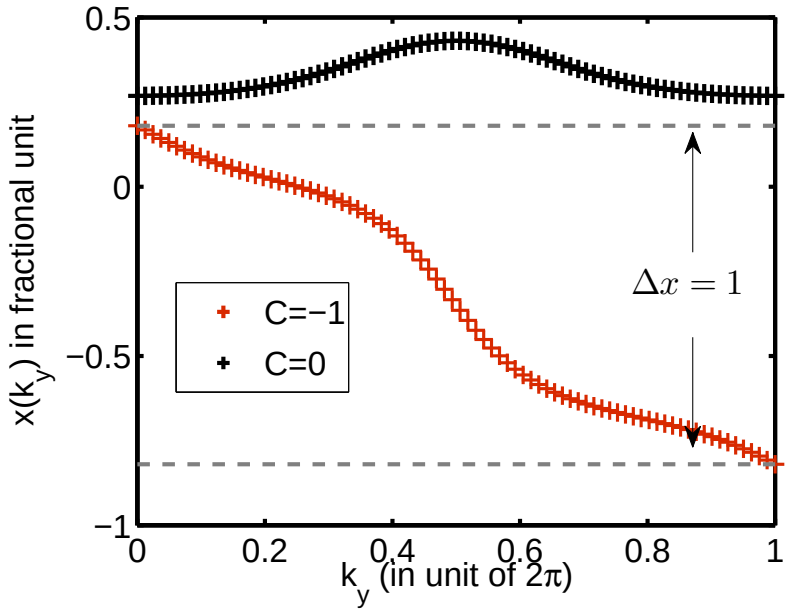

Figure 2.4: The hybrid Wannier center of the occupied band for the Haldane model ($x(k_y)$). The black and red crosses represent the normal phase and the Chern-insulator phase respectively.

edge states for a 2D Chern insulator with $C = \pm 1$. Such chiral edge states are nontrivial because the group velocities of the electrons staying in the edge states all point to one direction, thus the disorder-induced backscattering is not allowed near the edge. It should be emphasized again that such novel edge states stem from the bulk topological charge pump, which is protected by the bulk energy gap.

The Haldane model has motivated the search for Chern insulators in realistic materials. A crucial ingredient in the Haldane model is the breaking of $\mathcal{T}$ symmetry in the orbital degrees of freedom. However, in realistic systems, what usually happens is that $\mathcal{T}$ symmetry is broken spontaneously in the spin degrees of freedom, leading to spontaneous spin magnetizations. In order to break $\mathcal{T}$ symmetry in the orbital degrees of freedom, a non-negligible spin-orbit coupling (SOC) is needed. Therefore, what the Haldane model really describes is some simplified effective Hamiltonian of a quasi-2D system with some kind of spin magnetic order and non-negligible spin-orbit coupling.

Despite numerous proposals using both simplified model Hamiltonians and first

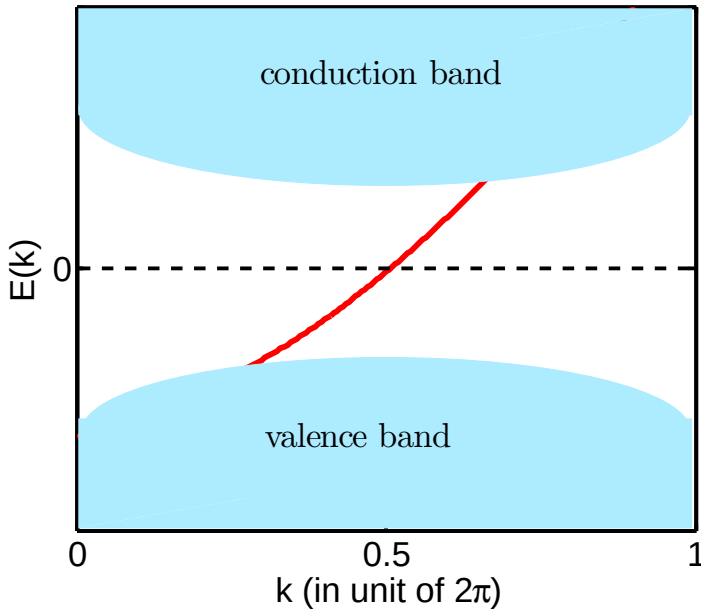

Figure 2.5: A schematic plot of the edge state in a Chern insulator with the magnitude of the Chern number to be 1.

principles calculations [83, 84, 85, 86], the Chern insulator remained as a theoretical hypothesis until it was experimentally confirmed in Cr-doped topological-insulator thin films in 2013 [25]. However, the size of the energy gap observed in Ref. [25] is on the order of meV, and the quantization of Hall conductance was observed only when the temperature dropped below ∼400 mK. Theorists keep coming up with new proposals, aiming to design an experimentally more robust QAH system. For example, one of the promising ideas is to deposite heavy adatoms with strong SOC such as Bi, on top of ferromagnetic thin films such as MnTe and MnSe [87]. It has predicted several promising candidates of Chern insulators with energy gaps on the order of 0.1 eV [87].

2.3.3 The quantum spin Hall effect

As discussed in the previous subsection, a Chern insulator can be characterized by a topological charge pump along the x direction as k_y varies adiabatically by 2π. After such an adiabatic cycle, an exactly integer number of charges are pumped from one edge to the other, where the integer is the Chern number defined in the (k_x, k_y) space. Let us

generalize the discussion to 2D insulating systems consisted of spin-1/2 fermions which respect time-reversal symmetry. As a result of $\mathcal{T}$ symmetry, the Chern number must vanish, corresponding to a vanishing net charge pump. However, if the spin degrees of freedom is taken into account, there could be a situation that the spin-up electrons and spin-down electrons moved toward opposite directions in response to the variation of k_y, resulted in a net spin pump and a non-vanishing spin Hall conductance for the 2D system.

In the same vein as the case of 1D charge pump, one can consider k_y as an external parameter that varies adiabatically and drives the nontrivial spin pump in a quasi-1D chain. As k_y increases, electrons with one spin species, say, spin up, would be pumped to the right end of the chain, going into an eigenstate localized at the right end whose energy level is raised up by the increase of k_y. In the meanwhile, the lowest unoccupied energy level of the same spin species localized at the left end would gradually drop down as k_y increases. On the other hand, the spin-down electrons are pumped towards the left end, giving rise to a spin-down eigenstate localized at the left end whose energy level increases with k_y. At some critical value of k_y, the unoccupied spin-up state going downward and the occupied spin-down state going upward (both are localized at the left end) would cross each other, then traverse through the bulk energy gap (Fig. 2.6). By virtue of the Kramers' theorem, the level crossing has to occur at either $k_y = 0$ or $k_y = \pi$, i.e., when time-reversal symmetry is preserved for the quasi-1D chain.

Now let us map those 1D chains parameterized by different k_y values back to a 2D system. If one plots the dispersions of the eigenenergies for a semi-infinite 2D system, one would see the gapless spin-polarized edge states as shown in Fig. 2.5. The two states localized at one edge possess opposite spin polarization and opposite momenta. Such novel edge states characterizes a new phase of electronic systems known as the quantum spin Hall (QSH) phase. These edge states are nontrivial because the net backscattering amplitude induced by any non-magnetic impurities near the edge vanishes as a result

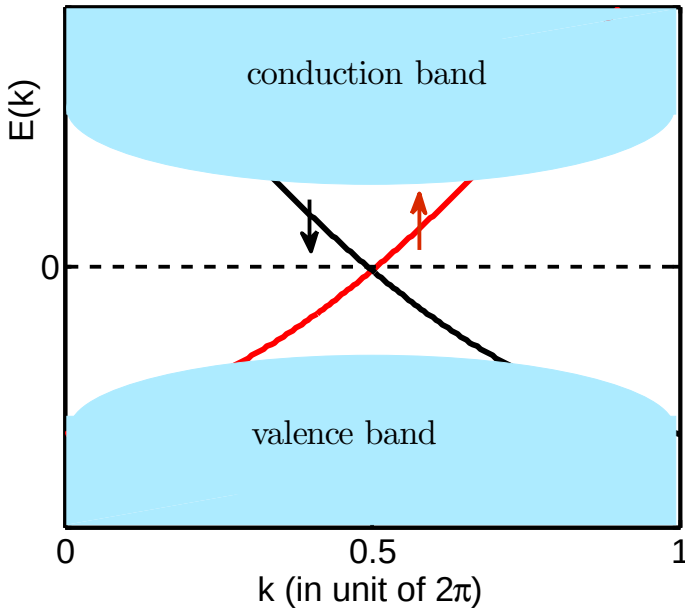

Figure 2.6: A schematic illustration of the spin-polarized edge states for the QSH insulator.

of time-reversal symmetry. Therefore, they are robust against Anderson localization induced by non-magnetic disorder. Electrons staying in these edge states would be in the regime of perfect ballistic transport, which would contribute a longitudinal charge conductance quantized as $2e^2/h$ (from the two edges) [88, 89].

The above discussion implicitly assumes the conservation of the spin quantum number s_z. When s_z is conserved, the net spin pumped after one cycle is quantized as $\hbar$, and the spin Hall conductance in the QSH phase is also exactly quantized as $e/2\pi$ [10]. However, in general there are spin-flip terms (e.g., Rashba spin-orbit coupling in 2D systems with breaking inversion symmetry) in the Hamiltonian so that there would be spin relaxations during the pumping process. Therefore, the net spin pumped after one cycle may be finite but not necessarily quantized. However, it turns out that the QSH phase remains robust in the presence of the spin-flip terms. It is even robust against weak disorder and weak electron-electron interactions as long as time-reversal symmetry is unbroken and the bulk energy gap is unclosed. This is because the QSH

[10] Without spin flip, the spin current density is defined as $\mathbf{J}_s = (\hbar/2e)(\mathbf{J}_\uparrow - \mathbf{J}_\downarrow)$, where $\mathbf{J}_\uparrow$ and $\mathbf{J}_\downarrow$ are charge currents from spin-up and spin-down electrons respectively.

phase is a topological phase of matter protected by time-reversal symmetry [26]. The nontrivial bulk band topology is characterized by the novel gapless edge states shown in Fig. 2.6.

To be more specific, the edge states in Fig. 2.6 are topologically in the sense that there is a single pair of spin-polarized gapless states traversing through the bulk energy gap, which cross each other at one of the TRIM. As mentioned above, backscattering with a spin flip is not allowed by $\mathcal{T}$ symmetry for such a single pair of edge states, therefore they are robust against nonmagnetic disorder. Moreover, as the level crossing at $k_y = 0$ or $k_y = \pi$ is protected by $\mathcal{T}$ symmetry, any perturbations that preserve $\mathcal{T}$ symmetry cannot lift the Kramers' degeneracy. In other words, with $\mathcal{T}$-invariant perturbations, the dispersions of the single pair of edge states may be modified and the position of the degeneracy point may be shifted up and down in energy, but they cannot be removed. Therefore, even if the spin-flip terms or other $\mathcal{T}$-invariant perturbations are included, the topological edge states would remain robust.

In principle there could be an arbitrary number of pairs of spin-polarized edge states which obey Kramers' degeneracy. However, unlike the case of Chern insulators, the topological classification in the presence of $\mathcal{T}$ symmetry is $\mathbb{Z}_2$ instead of $\mathbb{Z}$. This is because $\mathcal{T}$ symmetry only allows for the change of particle numbers by $2n$ for a given spin species [34]. It follows that backscattering is allowed only for an even number of pairs of the edge states. If the number of pairs of the edge states is odd, say, $2n+1$, then backscattering is allowed among $2n$ pairs of them, leaving one pair intact. Therefore there is a $\mathbb{Z}_2$ topological classification for 2D insulating systems with $\mathcal{T}$ symmetry.

In the remaining part of this subsection, we will review the first model Hamiltonian that realizes the QSH phase. We will calculate the hybrid Wannier centers of the model to better understand the bulk topological property. We conclude by brief discussions on the rigorous definition of the bulk $\mathbb{Z}_2$ index, and on the theoretical proposals and experimental realizations of the QSH phase in realistic material systems.

The quantum spin Hall effect was first proposed by Kane and Mele in 2005 [27]. They considered a 4-band tight-binding model on a 2D graphene lattice with time-reversal symmetry. The first-neighbor spin-independent hopping and the symmetry-allowed first-neighbor and second-neighbor SOC were considered in the model. The model Hamiltonian looks as

$$H = \sum_{\langle ij \rangle} t c_i^\dagger c_j + \sum_{\langle\langle ij \rangle\rangle} i\lambda_{\rm so} \nu_{ij} c_i^\dagger s_z c_j + \sum_{\langle ij \rangle} i\lambda_{\rm R} c_i^\dagger (\mathbf{s} \times \hat{\mathbf{d}}_{ij})_z c_j + \sum_i \epsilon (-1)^i c_i^\dagger c_i \,. \qquad (2.100)$$

where t is the first-neighbor spin-independent hopping amplitude, $\lambda_{\rm so}$ is the strength of the second-neighbor non-spin-flip SOC, $\lambda_{\rm R}$ is the first-neighbor Rashba-like SOC amplitude, and ϵ is the magnitude of the staggered on-site energies, with signs ± 1 for A and B sublattices respectively. Also, $\nu_{ij} = \pm 1$, with the sign depending on the chirality of the second-neighbor bond from site i to j, and $\hat{\mathbf{d}}_{ij}$ is the unit vector pointing from site i to its first-neighbor site j. In this model, $\lambda_{\rm so}$ competes with $\lambda_{\rm R}$ and ϵ, in the sense that $\lambda_{\rm so}$ tends to drive the system to the QSH phase while $\lambda_{\rm R}$ and ϵ tend to retain the trivial band topology. When s_z is conserved, the Kane-Mele model can be considered as a superposition of two copies of the Haldane model with opposite Chern numbers. If one calculates the 2D Chern numbers for spin-up and spin-down electrons separately, one would find that the two Chern numbers are ± 1 in the QSH phase.

In order to illustrate the bulk topological property, we calculate the hybrid Wannier centers $x(k_y)$ of the two occupied bands for the Kane-Mele model. As shown in Fig. 2.7(a), in the QSH phase, the two hybrid Wannier centers remain degenerate at $k_y = 0$ as a result of $\mathcal{T}$ symmetry. As k_y moves away from 0, the Kramer pair split apart and are separated by exactly one lattice constant along the x direction after a half cycle ($k_y = \pi$). If one plots multiple branches of the Wannier centers, it would be straightforward to see that these Kramers pairs switch partners between each other as k_y varies from 0 to π. As a result, if the 2D system is considered as a family of 1D chains parameterized by k_y, and truncations are made to these 1D chains, there would be unpaired excited charges of opposite spins localized at the two opposite ends when

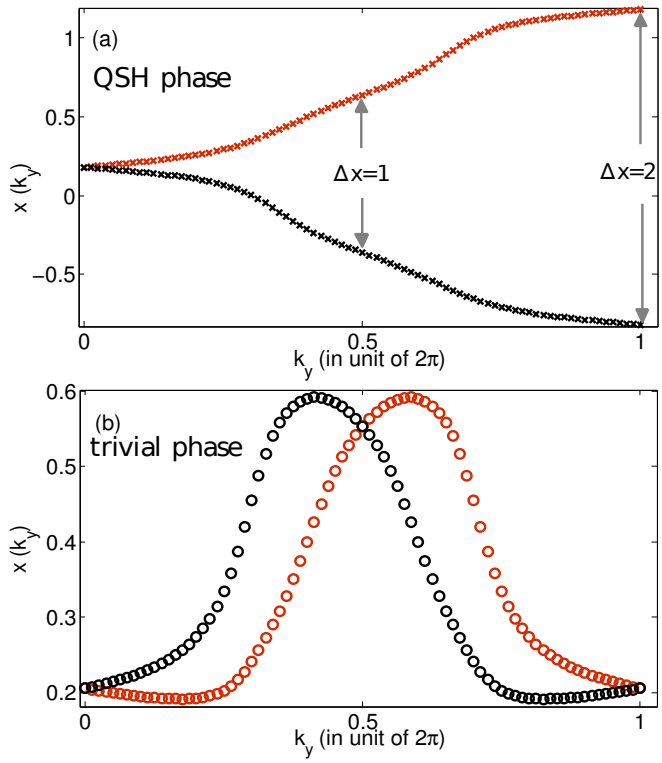

Figure 2.7: The hybrid Wannier centers of the two occupied bands for the Kane-Mele model ($x(k_y)$): (a) In the QSH phase (b) In the topologically trivial phase.

$k_y = \pi$. As k_y keeps increasing from π to 2π, the two Wannier centers continue to move away from each other. At $k_y = 2\pi$, they are split by exactly two lattice constants . Finally after a second cycle (not plotted), these two Wannier centers evolve back to the original state at $k_y = 0$.

On the other hand, in the topologically trivial phase, as shown in Fig. 2.7(b), the two Wannier centers split a little bit when k_y is away from 0, then become degenerate again at $k_y = \pi$. The Kramers' pairs never switch partners in the process of the adiabatic evolution. Therefore, there is no nontrivial edgestate, and the system is topologically trivial.

There are various ways to define the bulk $\mathbb{Z}_2$ index. The most intuitive approach is to associate the $\mathbb{Z}_2$ index to the "time-reversal polarization" P_x^θ along the x direction [56],

$$P_x^\theta(k_y) = P_x^I - P_x^{II}(k_y), \tag{2.101}$$

where $P_x^I(k_y)$ and $P_x^{II}(k_y)$ are the k_y-dependent 1D charge polarizations (Eq. (2.77)) for the two bands that form a Kramers' pair. They are exactly the Berry phases of the two bands calculated along the k_x direction at a given k_y (the electric charge e is set to 1),

$$P_x^{I(II)}(k_y) = \frac{i}{2\pi} \int_0^{2\pi} dk_x \, \langle u_{\mathbf{k}}^{I(II)} | \partial_x u_{\mathbf{k}}^{I(II)} \rangle, \tag{2.102}$$

where $\mathbf{k} = (k_x, k_y)$. The lattice constant along the x direction is normalized as 1. Actually the plots shown in Fig. 2.7 can be interpreted as the variation of P_x^I and P_x^{II} as a function of k_y.

The $\mathbb{Z}_2$ index ν_0 is defined as the difference between the time-reversal polarizations at $k_y = 0$ and $k_y = \pi$:

$$\nu_0 = P_x^\theta(k_y = \pi) - P_x^\theta(k_y = 0) \mod 2 \, . \tag{2.103}$$

When $\nu_0 = 1$, as shown in Fig. 2.7(a), the system is a QSH insulator; while when $\nu_0 = 0$ (Fig. 2.7(b)), the system is topologically trivial. It worth to note that the time-reversal polarization is a gauge-dependent quantity, but the difference between $P_x^\theta(k_y = \pi)$ and $P_x^\theta(k_y = 0)$ is a gauge-invariant. We have only considered a single Kramers' pair in Eq. (2.101)-(2.103), but the generalization to the case of N Kramers' pairs is straightforward.

When inversion symmetry is present in addition to time-reversal symmetry, the $\mathbb{Z}_2$ index can be determined by the parities of the occupied Kramers' doublets at the four TRIM [90] [11],

$$(-1)^{\nu_0} = \prod_{i=1}^{4} \prod_{n=1}^{N} \eta_n(\Gamma_i), \tag{2.104}$$

where $\eta_n(\Gamma_i)$ is the parity eigenvalue of the nth Kramers' pair of the Bloch states at the ith TRIM Γ_i, and N is the total number the occupied Kramers' pairs.

The QSH phase were theoretically predicted to exist in quite a few quasi-2D systems such as the HgTe/CdTe quantum wells [28], Bi thin films [91], silicene [92], and

[11] In 2D BZ, there are four TRIM, which are $(0,0)$, $(0,\pi)$, $(\pi,0)$ and (π,π).

numerous others [83, 93, 94]. As one of the earliest proposals of the QSH effect, the HgTe/CdTe quantum-well system is particularly interesting. To be specific, as the thickness of the quantum well (HgTe layer sandwiched between CdTe layers) increases, the system was predicted to go through a topological phase transition from a trivial to a QSH insulator. This proposal has been experimentally realized in HgTe/(Hg,Cd)Te quantum wells in 2007 [89].

2.3.4 3D Topological insulators

The definition of the 2D $\mathbb{Z}_2$ index can be generalized to 3D insulators with time-reversal symmetry. In the 3D BZ, there are six 2D planes which are "time-reversal invariant planes", i.e., the planes at $k_j = 0$ and $k_j = \pi$, with $j = 1, 2, 3$ labelling the directions of the three reciprocal lattice vectors. In principle one can define six 2D $\mathbb{Z}_2$ indices on these planes, namely, $\nu_j = \nu_{k_j=\pi}$ and $\nu'_j = \nu_{k_j=0}$, for $j = 1, 2, 3$. However, it turns out that only four of the six 2D $\mathbb{Z}_2$ indices are independent of each other. It is thus conventional to take the three indices ν_1, ν_2 and ν_3 as three weak topological indices, and $(-1)^{\nu_0} = (-1)^{\nu_1+\nu'_1} = (-1)^{\nu_2+\nu'_2} = (-1)^{\nu_3+\nu'_3}$ as the strong $\mathbb{Z}_2$ index [30, 32].

Since all the four indices $(\nu_0; \nu_1, \nu_2, \nu_3)$ are of $\mathbb{Z}_2$ type, in principle there are 16 classes of 3D insulators preserving $\mathcal{T}$ symmetry. These 16 classes of $\mathcal{T}$-invariant insulators generally fall into three categories: the trivial insulators, the weak topological insulators (TIs) and the strong TIs. The trivial insulators correspond to the situation that all the four indices are zeros. Most nonmagnetic insulators and semiconductors found in nature belong to this category. There is no nontrivial surface state in these topologically trivial materials. The weak topological insulators are characterized by a trivial strong index and at least one nontrivial weak index. The strong topological insulators are characterized by a nontrivial strong index, i.e., $\nu_0 = 1$.

The weak and strong TIs possess very interesting surface states. Let us first discuss the weak TIs. For example, if the weak index ν_1 is odd ($\nu_1 = 1$), then ν'_1 must also

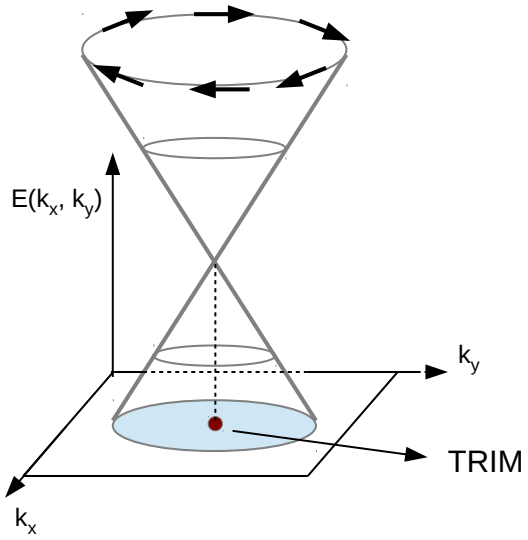

Figure 2.8: A schematic illustration of a surface Dirac cone centered at one of the TRIM of a projected surface BZ. The red point indicates the position of the TRIM. The black arrows represent the spin polarization at different wavevectors.

be odd because the strong index has to be even. Let us consider that a 3D system is consisted of a class of 2D systems parameterized by different k_1 values. Since $\nu_1 = 1$, the 2D system at $k_1 = \pi$ is a QSH insulator with an odd number of pairs of gapless edge states for any truncations to the 2D system. The same statement also applies to the 2D system at $k_1 = 0$, because both of them possess odd $\mathbb{Z}_2$ indices. As a result, if one make a surface of the 3D system whose surface normal is along either the k_2 or the k_3 direction, there would be an even number of gapless surface states centered at an even number of TRIM in the projected 2D surface BZ. As now we are in three dimensions, the gapless surface states form Dirac cones in the bulk energy gap with novel helical spin textures (Fig. 2.8(a)). On the other hand, if the surface normal is along the k_1 direction, there would not be any topological surface states.

For strong TIs, the strong $\mathbb{Z}_2$ index is odd $\nu_0 = 1$, which implies ν_j and ν'_j (for $j = 1, 2, 3$) must be opposite. Based on the argument on weak TIs, it follows that there must be an odd number of Dirac cones localized at the surface for arbitrary surface truncations. In the remaining part of this chapter, we will call the strong TIs as TIs for simplicity.

For 3D insulators with both time-reversal and inversion symmetry, it turns out that the $\mathbb{Z}_2$ index can be calculated from the parities of the occupied Bloch states at the eight TRIM in the 3D BZ. The formula for the strong $\mathbb{Z}_2$ index is simply the generalization of Eq. Eq. (2.104) to the case of eight TRIM,

$$(-1)^{\nu_0} = \prod_{i=1}^{8} \prod_{n=1}^{N} \eta_n(\Gamma_i), \tag{2.105}$$

where Γ_i now refers to the TRIM in the 3D BZ, and $\eta_n(\Gamma_i)$ is the parity of the nth Kramers' pair at Γ_i.

As mentioned in Sec. 2.2.4, a 3D TI is also characterized by a formally quantized Chern-Simons orbital magnetoelectric (ME) coupling $\alpha = e^2/2h$. As is well known, the time-reversal operation flips the sign of both magnetization and magnetic field $\mathbf{M} \to -\mathbf{M}$, $\mathbf{B} \to -\mathbf{B}$; whereas it leaves the electric polarization $\mathbf{P}$ and electric field $\mathbf{E}$ invariant. Therefore, for insulators with $\mathcal{T}$ symmetry, the magnetoelectric coupling coefficient α defined in Eq. (2.90) is expected to vanish. However, as discussed in Sec. 2.2.4, the Chern-Simons ME coupling α^{CS} is not unambiguously defined. The bulk electrodynamics of the media is invariant under $\alpha^{\mathrm{CS}} \to \alpha^{\mathrm{CS}} + e^2/h$, or equivalently $\theta \to \theta + 2\pi$ (Eq. (2.91)). Therefore there are two possible values of θ for systems with $\mathcal{T}$ symmetry, $\theta = 0$ or $\theta = \pi$. It turns out $\theta = \pi$ for TIs, while $\theta = 0$ for trivial insulators.

On the other hand, for a given sample of TI, the experimentally measured magnetoelectric coupling coefficient must be vanishing due to $\mathcal{T}$ symmetry. What happens is that the quantized θ coupling from the bulk is exactly cancelled by contributions from the metallic surface states, leading to a vanishing gross magnetoelectric response for the entire sample [95].

An interesting and challenging experiment is to gap the surface Dirac cones in TIs by "locally" breaking $\mathcal{T}$ symmetry only on (all) the surfaces of a TI sample, while the bulk still preserves $\mathcal{T}$ symmetry. Then the contribution to the ME coupling from the surface states vanish, and one would expect to obtain a quantized ME response contributed

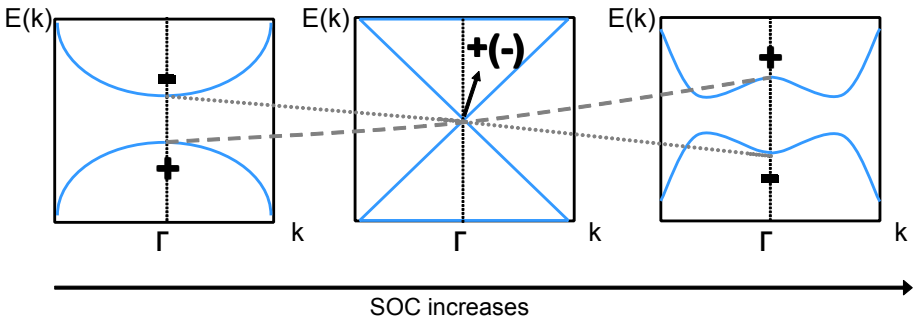

Figure 2.9: A schematic plot of the band-inversion process in Bi_2Se_3 at Γ as spin-orbit coupling increases.

purely by the bulk states. As discussed in Sec. 2.2.4, the isotropic bulk orbital ME coupling is equivalent to the anomalous Hall effect on the surface. Therefore, now the measured anomalous Hall conductance for the surface of the TI sample (with gapped surface states) would be "fractionally" quantized as $e^2/2h$ [34]. Unfortunately, such a intriguing thought experiment has not been realized yet due to technical difficulties.

Various materials have been theoretically proposed as 3D TIs, e.g., the Bi_2Se_3-class materials [36], some of the half-Heusler compounds [96], some of the ternary chalcogenides such as $TlBiSe_2$ [97], and numerous others as nicely summarized in Ref. [35]. Quite a few of these proposals have been experimentally verified [37, 38, 39, 98, 99]. Among these materials, the Bi_2Se_3-class TIs, including Bi_2Se_3, Bi_2Te_3 and Sb_2Te_3 [36], may be the most "popular" ones which have been extensively studied both theoretically and experimentally.

The Bi_2Se_3-class TIs have evoked significant interest in the community for several reasons. First of all, the physics picture in the Bi_2Se_3-class TIs is simple and elegant. Strong spin-orbit coupling (SOC) in Bi_2Se_3 would push up the highest occupied state and pull down the lowest unoccupied state at the center of the BZ (labeled as Γ). Since these two states possess opposite parities, they will cross each other at some critical SOC strength, then reopen a gap in an inverted order as shown in Fig. 2.9. As a result, the product of parities of all the occupied Kramers' pairs at Γ flips its sign, so does the strong $\mathbb{Z}_2$ index (Eq. (2.105)). Secondly, the "topological gap" after such "band

inversion" is quite large in the Bi_2Se_3-class TIs. For example, Bi_2Se_3 has a direct gap $\sim 0.5\,\mathrm{eV}$ at Γ and an indirect gap $\sim 0.3\,\mathrm{eV}$. It means that the topological properties such as the gapless surface states and the novel magnetoelectric response would remain robust up to room temperatures. Lastly, as the band inversion takes place only at Γ, there is only one surface Dirac cone in Bi_2Se_3 where the Dirac point is located at the center of the projected surface BZ. It is thus a perfect system to study the topological surface states.

On the other hand, experimentally the Bi_2Se_3-class TIs are not true insulators. For example, there are always Se vacancies in Bi_2Se_3 crystals, which dope electrons into the systems. Bi_2Te_3 is usually metallic due to the Bi-Te antisites. These extra charge carriers lift up (down) the Fermi level to the conduction (valence) band which hinders transport measurements on the topological surface states. However, significant progress has been made after years of consistent efforts in decreasing the bulk carrier densities [100, 101, 102, 103]. For example, by electric gating [100] and doping Ca [101], the bulk carrier in Bi_2Se_3 can be significantly reduced and transport signatures of the topological surface states have been successfully observed.

TIs also act as a platform for stimulating other new physics. For example, the Chern-insulator phase has been realized by doping magnetic impurities such as Cr into TI thin films [85, 25]. It was theoretically proposed that Majorana fermions could be realized at the surface of TIs by superconductor proximity effect [104]. In particular, a topological phase transition between a TI and a normal insulator may be realized by doping non-magnetic impurities into TIs. Depending on whether inversion symmetry is present or not, the critical phase may be either a 3D Dirac semimetal [105] or a topological Weyl semimetal [46, 48].

In the following two chapters, we will focus on the topological phase transitions in 3D TIs. We first study the topological phase transitions in In- and Sb-doped Bi_2Se_3, where the effects of disorder and the different orbital character introduced by impurities

are investigated. In Chapter 4, we deal with the topological phase transitions in TIs without inversion symmetry, and study the intermediate Weyl semimetals.

Chapter 3

Topological phase transitions in $(Bi_{1-x}In_x)_2Se_3$ and $(Bi_{1-x}Sb_x)_2Se_3$

In the previous chapter, we have discussed the concept of topological phases of matter and in particular 3D TIs. We mentioned that one of the most famous TIs is the Bi_2Se_3-class TIs, which has various merits as a type of promising topological materials. In this chapter, we study the phase transition from a topological to a normal insulator with concentration x in $(Bi_{1-x}In_x)_2Se_3$ and $(Bi_{1-x}Sb_x)_2Se_3$ in the Bi_2Se_3 crystal structure. We carry out first-principles calculations on small supercells, using this information to build Wannierized effective Hamiltonians for a more realistic treatment of disorder. Despite the fact that the spin-orbit coupling (SOC) strength is similar in In and Sb, we find that the critical concentration x_c is much smaller in $(Bi_{1-x}In_x)_2Se_3$ than in $(Bi_{1-x}Sb_x)_2Se_3$. For example, the direct supercell calculations suggest that x_c is below 12.5% and above 87.5% for the two alloys respectively. More accurate results are obtained from realistic disordered calculations, where the topological properties of the disordered systems are understood from a statistical point of view. Based on these calculations, x_c is around 17% for $(Bi_{1-x}In_x)_2Se_3$, but as high as 78%-83% for $(Bi_{1-x}Sb_x)_2Se_3$. In $(Bi_{1-x}Sb_x)_2Se_3$, we find that the phase transition is dominated by the decrease of SOC, with a crossover or "critical plateau" observed from around 78% to 83%. On the other hand, for $(Bi_{1-x}In_x)_2Se_3$, the In $5s$ orbitals suppress the topological band inversion at low impurity concentration, therefore accelerating the phase transition. In $(Bi_{1-x}In_x)_2Se_3$ we also find a tendency of In atoms to segregate.

3.1 Introduction

Topological aspects of quantum systems has been an exciting area in condensed-matter physics since the discovery of the integer quantum Hall effect (IQHE) [21, 22] and the first proposal of a 2D Chern insulator [24]. Both the IQHE and the 2D Chern insulators are characterized by a quantized Hall conductance and the presence of gapless edge modes that are topologically protected by a non-zero Chern number. In 2005, a topological classification was also found to apply to spinful systems with SOC and time-reversal symmetry, defining a topologically non-trivial 2D state known as a quantum spin Hall (QSH) insulator [27, 26]. A QSH insulator also possesses gapless edge states that always cross at one of the time-reversal invariant momenta (TRIM) in the 1D edge Brillouin zone (BZ). In a 2D Chern insulator, the chiral gapless edge modes can be interpreted in terms of the charge accumulation at one end of a truncated 1D system during an adiabatic periodic evolution. The spin-polarized edge modes in a QSH insulator can be interpreted in a similar way, except that charges with opposite spin characters are pumped in opposite directions and accumulated on opposite ends [56]. This pumping process can be classified by a new topological index, known as the $\mathbb{Z}_2$ index, which guarantees the robustness of the edge modes of a QSH insulator to weak time-reversal invariant perturbations.

The definition of the $\mathbb{Z}_2$ index was later generalized from 2D to 3D crystals [30, 106]. In 3D systems, there is one strong $\mathbb{Z}_2$ index, which is odd when the number of Dirac cones on the surface is odd, defining a strong topological insulator (TI); when it is even, the other three indices characterize the weak TIs, specifying how these gapless surface states are distributed among the TRIM in the 2D surface BZ.

A non-trivial bulk topological index is usually connected with a non-trivial "topological gap" resulting from band inversion. For systems with inversion symmetry, the topological index can be uniquely determined from the parities of the occupied states

at the TRIM in the BZ [90]. Thus, to drive an inversion-symmetric system from a normal insulator (NI) to a TI, a strong SOC is usually needed to flip the valence-band maximum (VBM) and conduction-band minimum (CBM) with opposite parities at one of the TRIM. The band gap after the topological band inversion is conventionally assigned with a minus sign, to be distinguished from the ordinary band gap in the $\mathbb{Z}_2$-even case. The scenario sketched above is exactly the mechanism in the Bi_2Se_3 class of TIs [36, 107, 38, 37, 108, 109, 110, 39]. In Bi_2Se_3 with SOC turned off, the VBM and CBM states at Γ are built from Se $4p$ and Bi $6p$ orbitals in such a way as to have opposite parities. When SOC is turned on, the previous VBM is pushed up into the conduction bands, leading to an exchange of parities and a non-trivial $\mathbb{Z}_2$ index. As long as the inverted band gap remains and time-reversal symmetry is preserved, a single Dirac cone exhibiting a helical spin texture is guaranteed to exist at $\overline{\Gamma}$ in the surface BZ. For some useful recent reviews, see Refs. [33, 34, 111].

Up to now, however, only a few pioneering works [112, 113, 114] have focused on the topological phase transition from the TI to the NI state driven by non-magnetic substitution, and while the general picture of such a transition seems obvious, details remain unclear. In the simplest picture, one would expect the band gap of a TI to decrease linearly as a lighter element with weaker SOC is substituted, and the phase transition would occur when the bulk gap is closed. However, on a closer look, many questions arise. For example, the bandstructures of known TIs are mostly dominated by p orbitals, but what happens if the substituted element includes different valence orbitals such as s or d orbitals? More fundamentally, translational symmetry is lost for a randomly substituted system. In this case, how should one determine the topological properties of a system in which wavevector $\mathbf{k}$ is no longer a good quantum number, and what signature indicates the presence of a TI state? These questions focus on two aspects that are not taken into account in the simplest linear band-closure picture: the effects of impurities with different orbital character, and the effects of disorder.

These issues arise, in particular, for the substitution of In into Bi_2Se_3, one of the best-known TI systems. Recently, several experimental groups have reported a surprisingly low critical concentration x_c of about 5% in $(Bi_{1-x}In_x)_2Se_3$, much lower than would be expected from a linear band-closure picture, thus challenging the usual understanding of the phase-transition behavior of TIs by non-magnetic doping [41, 40]. These experiments motivated our theoretical studies of the $(Bi_{1-x}In_x)_2Se_3$ system. Moreover, to separate the effects of In $5s$ orbitals from a simple weakening of the effective SOC, we also study $(Bi_{1-x}Sb_x)_2Se_3$. Here Sb has the same orbital character as Bi, lying directly above it in the Periodic Table, but shares the weaker intrinsic SOC strength of In because their atomic numbers are very close in magnitude.

We first study the solid-solution systems by constructing small supercells with different impurity configurations. For each supercell configuration, the strong $\mathbb{Z}_2$ index and surface states are computed using Wannier-interpolation techniques [9], which also allow us to test the effect of artificially removing the In $5s$ orbitals from the calculation. Next, we study the the effects of disorder more realistically by constructing a large supercell of pure Bi_2Se_3 acting as reference system, making random substitutions of In or Sb on the Bi sites, and calculating the disorder-averaged spectral functions [115, 116]. We further propose an approach in which we compute "$\mathbb{Z}_2$-index statistics" in order to determine the topological properties of disordered systems from a statistical point of view.

Based on our results, the $(Bi_{1-x}Sb_x)_2Se_3$ system is well described by the linear band-closure picture with a high critical concentration x_c, because the orbital character of the host and dopant are the same and the disorder effect is thus rather weak. We also observe a "critical plateau" in the Sb-substituted system, where the critical Dirac semimetal phase remains robust from about $x \approx 78\%$ to $x \approx 83\%$, although it is difficult to test whether this may be a finite-size effect due to the limited numerical accuracy in our calculations. In In-substituted Bi_2Se_3, on the other hand, the disorder effects are

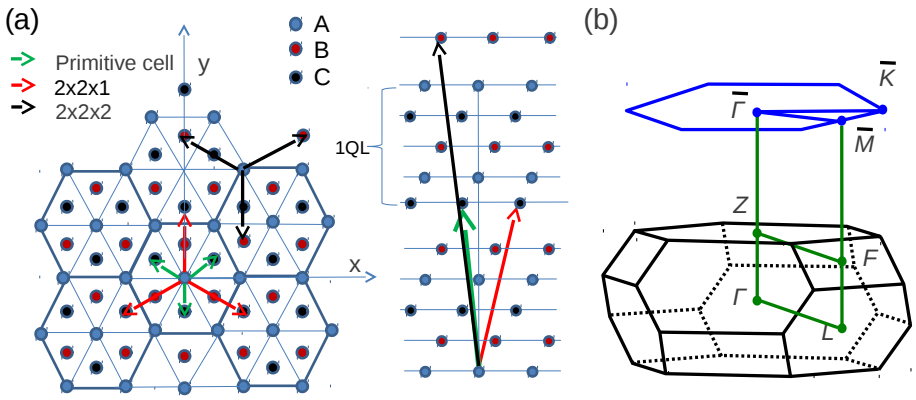

Figure 3.1: (a) Lattice vectors of Bi_2Se_3 primitive cell, 2×2×1 supercell, and 2×2×2 supercell. (b) Corresponding bulk Brillouin zone and its surface projection.

quite strong, and the presence of In $5s$ orbitals rapidly drives the system into the NI state even at very low impurity concentrations. A tendency of segregation of In atoms has been observed for $(Bi_{1-x}In_x)_2Se_3$, and may play an important role.

This chapter is organized as follows. In Sec. II the lattice structures, notations and the details of first-principles calculations are introduced. In Sec. III we present the main results of this chapter. First we summarize the results from the direct first-principles superlattice calculations, and determine the critical points and the influence of In $5s$ orbitals by computing the bulk $\mathbb{Z}_2$ index and by calculating surface states. Then, the critical points of the two solid-solution systems are further determined by looking at the disordered spectral functions, and the topological behaviors are understood from a statistical point of view. Finally we summarize in Sec. IV.

3.2 Preliminaries

3.2.1 Structures of bulk material and superlattices

As shown in Fig. 3.1, the crystal structure of Bi_2Se_3 is rhombohedral. The crystal has a layered structure along the z direction with five atoms per primitive cell. The five 2D monolayers made by repeating the primitive cell in the x and y directions form a quintuple layer (QL). In each QL, there are two equivalent Se atoms located at the

top and bottom of the QL, two equivalent Bi atoms inside those, and another central Se atom. Seen from the top, each monolayer forms a 2D triangular lattice, and these triangular planes are stacked along the z direction in the order $A-B-C-A-B-C...$, where A, B and C represent the three different high-symmetry sites. Both Bi_2Se_3 and β-phase In_2Se_3 have a rhombohedral structure belonging to the $R\overline{3}$m space group, but their lattice parameters are slightly different. The in-plane hexagonal lattice parameter is $a = 4.138$ Å for Bi_2Se_3 but $a = 4.05$ Å for In_2Se_3, and the height of a QL is $c = 9.547$ Å for Bi_2Se_3 compared with $c = 9.803$ Å for In_2Se_3. The rhombohedral structure of Sb_2Se_3 does not exist in nature, so for this case we relaxed both the lattice parameters and atomic positions. After a complete relaxation, we obtained $a = 4.11$ Å and $c = 10.43$ Å for Sb_2Se_3.

To study the substitution problem from first-principles calculations, a $2\times2\times1$ supercell based on the original Bi_2Se_3 crystal structure is built. The lattice vectors of the supercell are shown in Fig. 3.1. There are 20 atomic sites in such a supercell, where eight of them are Bi sites. Among all the possible configurations, we choose to investigate the supercells with one, two, four, six and seven Bi atoms substituted by impurities. The (unique) configuration with $x = 0.125$ is denoted as $C_{0.125}$. For two or six impurities ($x = 0.25$ and $x = 0.75$), there are two inequivalent configurations, with the two impurity (or remaining host) atoms residing in different monolayers or in the same monolayer, which we label as $C_{0.25}$ ($C_{0.75}$) and $C'_{0.25}$ ($C'_{0.75}$) respectively. For four impurities, $x = 0.5$, all impurities can be clustered in one monolayer, labeled as $C''_{0.5}$, or three in one monolayer and one in the other, denoted as $C'_{0.5}$, or the impurities can be equally divided between top and bottom monolayers with inversion symmetry, denoted as $C_{0.5}$. Note that primes indicate more strongly clustered configurations.

3.2.2 First-principles methodology

The first-principles calculations are carried out with the QUANTUM ESPRESSO package [117], with the PBE generalized gradient approximation (GGA) exchange-correlation functional [118, 119] and well-tested fully relativistic ultrasoft [120] and norm-conserving pseudopotentials. The ultrasoft pseudopotentials are from QUANTUM ESPRESSO [1], and the norm-conserving pseudopotentials are from the OPIUM package [121, 122]. The ionic relaxations, ground-state energies, and densities of states presented in Sec. 3.3 are calculated with ultrasoft pseudopotentials, but we switched to norm-conserving pseudopotentials for those topics that required transformation to a Wannier representation (see below). The energy cutoff with ultrasoft pseudopotentials is 60 Ry for In-substituted Bi_2Se_3 supercells and 35 Ry for Sb-substituted supercells. The cutoff becomes larger for norm-conserving pseudopotentials, specifically 65 Ry for In substitution and 55 Ry for Sb substitution. The BZ is sampled on a 6×6×6 Monkhorst-Pack [123] **k** mesh for the 2×2×1 supercells, and 8×8×8 for the primitive cell bulk materials. In our calculations, the lattice parameters of the Sb- and In-substituted supercells are fixed, taken as a linear interpolation of the Bi_2Se_3 and In_2Se_3 experimental lattice parameters according to the impurity concentration x, and the internal coordinates of the atoms are fully relaxed. We do not relax the lattice vectors because the coupling between two QLs is at least partially of van der Waals type, so that the standard GGA does not give a good estimate of the lattice constants, especially the one in the z direction.

To investigate the topological properties of these supercells, we calculate both the bulk $\mathbb{Z}_2$ indices and the surface states using the Wannier-interpolation technique. More specifically, we use the Wannier90 package to generate Wannier functions (WFs) from the outputs of standard first-principles calculations [124]. Wannier90 can optionally generate maximally localized WFs [52, 53], and in any case reports the Wannier charge

[1] We used the pseudopotentials Bi.rel-pbe-dn-rrkjus.UPF, Se.rel-pbe-n-rrkjus.UPF, Sb.rel-pbe.US.UPF and In.rel-pbe-dn-rrkjus.UPF from http://quantum-espresso.org

centers, their spreads, and the real-space Hamiltonian matrix elements of an effective tight-binding (TB) model in the WF basis. This information is often very useful in studying the bonding mechanism of materials, as well as for calculating topological indices, computing surface and interface states, treating disorder, etc.

It should be noted that the TB models constructed from Wannier90 are realistic in the sense that the Wannier-interpolated bandstructures reproduce the first-principle bandstructures essentially exactly within a certain energy window. This "frozen window" is chosen to extend from 3 eV below the Fermi level to 3 eV above the Fermi level in our calculations. In addition to the frozen window, there is also an outer energy window outside which the Bloch eigenstates will not be included in generating the WFs. The outer window varies in our calculations depending on the system, but typically covers a total range of 17-22 eV and includes all the valence p bands as well as In valence s bands when present. For example, for Bi_2Se_3we construct 30 spinor WFs per primitive cell, and two additional WFs constructed from In valence s orbitals would be added for each substituted In atom.

3.3 Results and discussions

3.3.1 Ground-state energies and band gaps

We begin by discussing our results for In-substituted supercells representing $(Bi_{1-x}In_x)_2Se_3$. The ground-state energies for supercells with different In impurity configurations are shown in Fig. 3.2. Open and closed circles represent topologically normal and $\mathbb{Z}_2$-odd cases respectively (see Sec. 3.3.4). For concentrations $0.25 \leq x \leq 0.75$ there are two or more inequivalent configurations of the $2\times2\times1$ supercell having the same concentration x. Among these, the configurations with lowest total energy are traced by the solid red line, and are found to consist of "clustered" configurations in which the In impurities tend to be first neighbors. Conversely, those with the highest total energies,

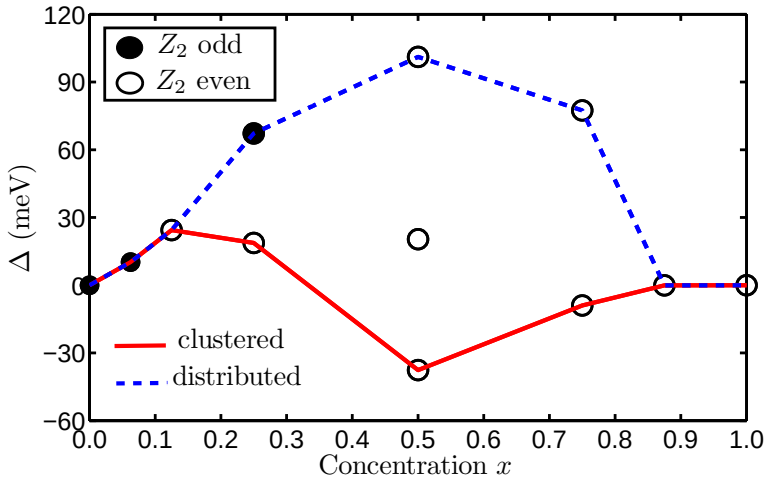

Figure 3.2: Ground-state energies vs. impurity concentration x for $(Bi_{1-x}In_x)_2Se_3$ supercells. Here $\Delta = E_g(x) - (1-x)E_1 + xE_2$, where $E_g(x)$, E_1 and E_2 are the ground-state energies per 5-atom cell for the alloy supercell, host material, and dopant material, respectively. Filled and open circles denote $\mathbb{Z}_2$-odd and even states respectively. Solid (red) and dashed (dark blue) lines follow the most and least In-clustered configurations respectively.

indicated by the dark blue dashed line, are those with the In atoms distributed most evenly throughout the supercell. For example, at $x = 0.5$, the ground-state energy of the clustered configuration ($C''_{0.5}$) is lower than that of the distributed one ($C_{0.5}$) by 140 meV per primitive unit cell, and at $x = 0.25$ the energy of $C'_{0.25}$ is lower than that of $C_{0.25}$ by 50 meV per primitive unit cell. Thus we clearly find a strong tendency of the In atoms to segregate and cluster together. We also find that the $\mathbb{Z}_2$ index changes sign at a critical concentration x_c lying somewhere between 6.25% and 12.5%. (One may notice from Fig. 3.2 that the distributed configuration $C_{0.25}$ at $x = 0.25$ is $\mathbb{Z}_2$-odd, but since its energy is so much higher, the significance of this is questionable.)

Turning now to the case of Sb substitution, we find a quite different behavior. The corresponding total-energy results for the $2\times2\times1$ $(Bi_{1-x}Sb_x)_2Se_3$ supercells are presented in Fig. 3.3(a). Here we find that the energies of different configurations at the same x differ by no more than 10 meV per primitive unit cell, which is roughly ten times smaller than in $(Bi_{1-x}In_x)_2Se_3$ (note the difference in the vertical scales here compared to Fig. 3.2). This signifies that the disorder effect is very weak in this system.

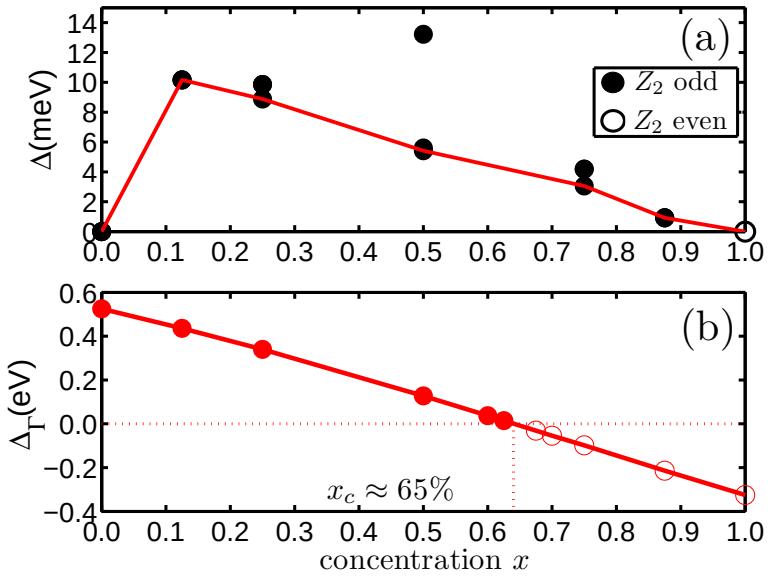

Figure 3.3: (a) Ground-state energies vs. impurity concentration x for $(Bi_{1-x}Sb_x)_2Se_3$ supercells, following the same conventions as in Fig. 3.2. (b) Band gap at the center of the Brillouin zone vs. impurity concentration x computed within the virtual crystal approximation. Positive and negative values of band gap denote the topological and normal phases respectively.

It is also evident from Fig. 3.3(a) that the system remains in the TI phase even up to $x = 87.5\%$, in sharp contrast to the behavior in $(Bi_{1-x}In_x)_2Se_3$.

Because we find the disorder effect to be so weak in Sb_2Se_3, we have also analyzed its behavior using the virtual crystal approximation (VCA), in which each Bi or Sb is replaced by an identical average atom whose properties are a weighted mean of the two constituents. We implement the VCA in a Wannier basis by constructing separate 30-band models for Bi_2Se_3 and Sb_2Se_3, including all the valence cation and anion p orbitals. The Hamiltonian matrix elements H_{mn}^{VCA} of the "virtual crystal" are taken as the linear interpolation in x of the two bulk materials, $H_{mn}^{\mathrm{VCA}} = (1-x)H_{mn}^{\mathrm{Bi}} + xH_{mn}^{\mathrm{Sb}}$, where H_{mn}^{Bi} and H_{mn}^{Sb} denote the matrix elements of the TB models of Bi_2Se_3 and Sb_2Se_3. We note in passing that one has to be cautious when generating the WFs for the VCA procedure, since it is important for the Wannier basis functions to be as similar as possible before the averaging takes place. Only in this way will the addition and subtraction between two different Hamiltonians be well defined. Because the maximal

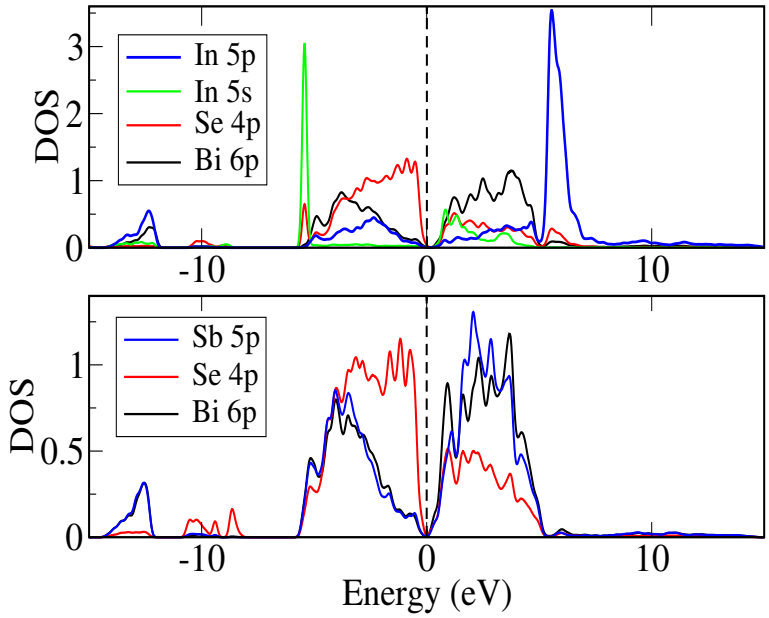

Figure 3.4: (a) Local DOS of $(Bi_{1-x}In_x)_2Se_3$ at $x = 12.5\%$ for the s and p orbitals on the substituted In atom and the p orbitals on first-neighbor Bi and Se atoms. (b) Local DOS of $(Bi_{1-x}Sb_x)_2Se_3$ at $x = 12.5\%$ for the p orbitals on the substituted Sb atom and the p orbitals on first-neighbor Bi and Se atoms.

localization procedure might generate different WFs for different systems as it seeks to minimize the "spread functional" [52], we construct the WFs for the VCA treatment simply by projecting the Bloch states onto the same set of atomic-like trial orbitals without any further iterative localization procedure.

Within this VCA approach, it is straightforward to compute the band gaps and topological indices, since only a primitive bulk cell is needed. Fig. 3.3(b) shows how Δ_Γ, the band gap at the Brillouin zone center, evolves with x for the $(Bi_{1-x}Sb_x)_2Se_3$ virtual crystal. It is evident that the gap closes at $x_c \simeq 65\%$, where the system undergoes a transition to the normal-insulator state (here indicated by a negative gap value).

3.3.2 Orbital character

To get some physical insight about the distinct behaviors in the two substituted systems, we turn to study the orbital character at a low composition of $x = 12.5\%$. The local density of states (DOS) of the substituted In and Sb atoms and their neighboring Bi

and Se atoms are plotted in Fig. 3.4. For low-composition In-substituted systems, the In $5s$ orbitals and nearest-neighbor Se $4p$ orbitals form bonding and antibonding states, with the former leading to a flat band deep in the valence bands corresponding to the In $5s$ peak around $-6\,\mathrm{eV}$ in Fig. 3.4(a). The hybridized s-p antibonding states further interact with the Bi $6p$ orbitals, bringing some In $5s$ character into the conduction bands. The In $5p$ orbitals are mainly responsible for the sharp peak about $7\,\mathrm{eV}$ above the Fermi level in Fig. 3.4(a), but also mix with Bi and Se p orbitals on the nearby atoms to contribute to the lower conduction-band states. The hopping between In $5p$ and neighboring Se $4p$ states, on the other hand, contributes mainly to the valence band, but also to the lower conduction bands.

If one only focuses on the low-energy physics, say within $5\,\mathrm{eV}$ of the Fermi level, one would notice that the In $5p$ states are homogeneously distributed among the valence and conduction bands. On the other hand, the s orbitals are more concentrated at the bottom of the valence and conduction bands. This implies that the effects of In $5s$ and $5p$ orbitals in the supercell electronic structure are distinct. The non-homogeneously distributed In $5s$ states may be crucial in determining the topological properties of the supercell. From the DOS at the Γ point (not shown here), we also observe that the VBM is mostly composed of Se $4p$ states, while the CBM is dominated by Bi $6p$ states. This implies that the nontrivial topological band inversion has already been removed at 12.5% of In substitution.

For the Sb substitution at $x = 12.5\%$, however, the local DOS shown in Fig. 3.4(b) indicates that that Sb $5p$ orbitals are more or less homogeneously distributed among the valence and conduction bands as they hybridize with the Bi and Se p states. In fact, the Sb 5p and Bi 6p local DOS profiles are strikingly similar. While not displayed, we also explore the DOS of $(Bi_{1-x}Sb_x)_2Se_3$ at other compositions, and observe that the hybridization between Bi, Se and Sb p states remains homogeneous over the entire composition range. A homogeneous hybridization of Bi, Se and Sb p states tends to

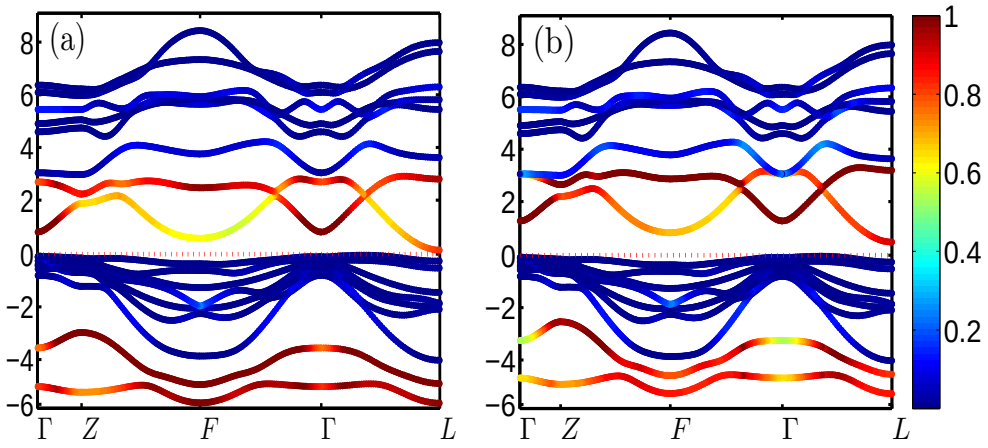

Figure 3.5: (a) Wannier-interpolated bandstructure of In_2Se_3, with color code indicating In $5s$ character. (b) Same but with In $5s$ levels shifted upward by 0.79 eV.

confirm the appropriateness of the use of VCA with artificial orbitals to construct an effective description of the electronic structure of the substituted system. Within the VCA, the strength of the effective SOC would be expected to decrease linearly as x is increased, suggesting that the topological phase transition in the Sb-substituted system should belong to the linear band-closure regime.

To study In-substituted Bi_2Se_3 at a lower concentration, a 2×2×2 supercell has been constructed, in which one out of 16 Bi atoms is substituted by In. The supercell lattice vectors are shown in Fig. 3.1. The energy cutoff is taken to be the same as for the 2×2×1 supercell calculations. A 3×3×3 Monkhorst-Pack **k** mesh is used for ionic relaxation and calculation of the ground-state energy, while it is increased to 4×4×4 for the non-self-consistent calculation used to interface with Wannier90. The ground-state energy is indicated by the filled circle at $x = 6.25\%$ in Fig. 3.2, and is confirmed to be in the $\mathbb{Z}_2$-odd phase from the $\mathbb{Z}_2$ index and surface-state calculations.

3.3.3 Shift of In $5s$ levels

Among the In-substituted configurations, our calculations find that $C_{0.5}$ and $C_{0.75}$ are metallic, in contrast with experimental observations showing the gap opening with increasing x beyond the transition to the normal phase [41, 40]. The reason for the gap

closure becomes clear from an inspection of our calculated bandstructure of bulk In_2Se_3, shown in Fig. 3.5(a), which was computed using the Wannier interpolation capabilities of the Wannier90 package [124] based on a 34-band TB model including the 30 valence p orbitals and four In $5s$ orbitals. The color coding in Fig. 3.5(a) shows the degree of In $5s$ character. We find that there is almost a gap closure, $E_g \simeq 0.15\,\text{eV}$, much smaller than the experimental value of 1.34 eV [125]. Our small gap clearly results from a low-lying conduction band at L that is dominantly of In $5s$ character. For the $C_{0.5}$ and $C_{0.75}$ cases, these states get folded and mixed with other conduction-band states in such a way as to cause the metallic behavior observed in our supercell calculations.

We have good reason to believe, however, that the energy position of these In $5s$ is incorrectly predicted by standard density functional theory (DFT) [3, 4]. It is well known that DFT tends to underestimate gaps, especially when the character of the VBM (here p states at Γ) and the CBM (here In $5s$ states at L) are different. More specifically, however, quasiparticle calculations on InAs have shown that the In $5s$ energy levels are too low relative to the many-body GW calculation [126]. In particular, the CBM at Γ, having In $5s$ character, was found there to be too low by about 0.79 eV within DFT. We have checked that our In $5s$ energy positions do not depend sensitively on the use of the local-density approximation (LDA) [127] *vs.* GGA, the choice of pseudopotentials, or the use of different code packages [128, 129]. Therefore, we conclude that more advanced approaches such as hybrid functionals or direct many-body methods are needed to fix this problem.

Unfortunately, application of hybrid functionals to our supercell calculations would be computationally expensive. Here we have taken a simpler approach to adjust the In $5s$ energy levels. The Wannier interpolation procedure has already provided us with a first-principles effective TB model reproducing the DFT bandstructure. We simply shift the energies of all the In $5s$ orbitals within this effective model upward by 0.79 eV, the value taken from Refs. [126], and leave all other matrix elements unchanged. The

resulting bandstructure for bulk In_2Se_3 is shown in Fig. 3.5(b). We find that the band gap opens up to 0.52 eV, while otherwise the general character of bandstructure is not significantly changed.

While 0.52 eV is still far from an experimentally correct estimate of the gap, we expect our modified Wannier Hamiltonian should be good enough for the purpose of computing topological properties of $(Bi_{1-x}In_x)_2Se_3$ solid solutions. Once we apply this shift, we find that that the supercells that were metallic before are now insulating, and moreover the states near the Fermi energy that determine the topological character do not have significant In $5s$ character. Therefore, the magnitude of the shift is not important for computing the topological properties, as long as it is large enough to prevent the In $5s$ levels from interfering. In any case, since the $\beta-$ phase of In_2Se_3 is not very stable at room temperature (it has to be stabilized by doping small amounts of Sb) [125], a direct comparison between the experimental and theoretical band gaps is not very meaningful. Therefore, we adopt the procedure here of applying the 0.79 eV shift of In $5s$ levels in all of our In-substituted supercell calculations. In particular, the $\mathbb{Z}_2$ indices (filled *vs.* open circles) shown in Fig. 3.2 have been computed using this shift, as will be discussed in detail next.

3.3.4 $\mathbb{Z}_2$ indices

The strong $\mathbb{Z}_2$ indices of all the In and Sb-substituted Bi_2Se_3 supercells have been calculated in order to locate the critical concentrations for the transition from topological to normal behavior in the two solid-solution systems. As discussed above, some of the supercells ($C_{0.25}$, $C_{0.5}$ and $C_{0.75}$) have inversion symmetry, in which case the strong $\mathbb{Z}_2$ index can be evaluated simply by counting the parities of the occupied bands at the TRIM in the BZ. Specifically, if one defines δ_i as the product of the parities of the occupied bands (counting just one band from each Kramers doublet) at the ith TRIM in the BZ, the strong $\mathbb{Z}_2$ index is just $\nu_0 = \prod_{i=1}^{8} \delta_i$, i.e., the product of δ_i at all the

eight TRIM [90].

In the general case, however, the strong $\mathbb{Z}_2$ index has to be determined by explicitly calculating the 2D $\mathbb{Z}_2$ indices of the top and bottom slices of half of the 3D BZ. There are six such 2D indices, namely $\nu_j \equiv \nu_{k_j=0}$ and $\nu'_j \equiv \nu_{k_j=\pi}$, corresponding to the indices of the slices at $k_j = 0$ and $k_j = \pi$, where $j = \{1, 2, 3\}$ labels the three wavevector directions in the BZ. However, only four of the six indices turn out to be independent variables. The indices ν_1, ν_2 and ν_3 are usually taken to define the three weak topological indices, while the product $\nu_0 = \nu_j \nu'_j$ of the two indices on any pair of parallel slices is known as the strong $\mathbb{Z}_2$ index ν_0. This means that if two parallel slices have different $\mathbb{Z}_2$ indices, as for example at $k_3 = 0$ and $k_3 = \pi$, then ν_0 is odd and the system is a strong TI; otherwise it is a weak TI if any indices are odd, or normal if not.

In the absence of inversion symmetry, the 2D $\mathbb{Z}_2$ index is defined by the change of 1D "time-reversal polarization," say in the k_1 direction, as the other wavevector k_2 evolves from 0 to π. The time-reversal polarization can be explicitly visualized by tracing the 1D hybrid Wannier charge centers (WCCs) [52] in the k_1 direction as a function of k_2. The $\mathbb{Z}_2$ index is odd if the hybrid WCCs of the Kramers doublets switch partners during the evolution, and even otherwise [56].

We implement these ideas in practice using the approach of Soluyanov and Vanderbilt [130], in which the 2D $\mathbb{Z}_2$ index is obtained by counting the number of jumps of the "biggest gap" among the 1D hybrid WCCs during the evolution. The approach for computing the $\mathbb{Z}_2$ indices described above has been implemented in the Wannier basis using the matrix elements of the effective TB model and the WCCs generated from Wannier90 [124].

The results are shown in Figs. 3.2 and 3.3(a) by using filled circles to indicate cases in which the strong Z2 index is odd, while an open circle means it is even. In fact, none of the $\mathbb{Z}_2$-even configurations are found to be weak TIs, so open circles denote topologically normal insulators. If one follows the solid (red) lines in Fig. 3.2 and Fig. 3.3(a), which

Table 3.1: Bulk band gaps at Γ for $x = 12.5\%$ in $(Bi_{1-x}In_x)_2Se_3$. "Shifted" In $5s$ levels were raised by 0.79 eV (see text).

	without SOC (eV)	with SOC (eV)
In $5s$ levels unshifted	0.72	0.11
In $5s$ levels shifted	0.68	0.07
In $5s$ levels removed	0.42	−0.26

track the configurations with lowest energies, it is clear that for $(Bi_{1-x}In_x)_2Se_3$ the system becomes topologically trivial for $x > 6.25\%$. For $(Bi_{1-x}Sb_x)_2Se_3$, however, the TI phase is preserved up to 87.5%.

It should be emphasized again that a 0.79 eV shift has been added on the on-site In $5s$ energy levels in the effective TB models for the supercells of $(Bi_{1-x}In_x)_2Se_3$. However, except for $C_{0.5}$ and $C_{0.75}$, which are metallic without the shift, the $\mathbb{Z}_2$ indices of all the other configurations are unchanged by the application of this shift.

3.3.5 Effects of In $5s$ orbitals on bulk and surface states

To understand why the phase transition happens so rapidly in $(Bi_{1-x}In_x)_2Se_3$, we focus on the $x = 0.125$ supercell, and separately investigate the effects of In $5s$ orbitals and SOC on bulk bandstructure. As shown in Table 3.1, without SOC the In $5s$ orbitals try to pull down the VBM, leading to a band gap as large as 0.7 eV at Γ, such that the SOC strength is not large enough to invert the CBM and VBM. If the In $5s$ orbitals are removed, however, the gap at Γ is only 0.42 eV without SOC, and when SOC is added back the band inversion reoccurs, with an inverted gap as large as 0.26 eV (denoted with a minus sign in Table 3.1). We also notice that the shift of In $5s$ levels only changes the gap at Γ by 0.04 eV, and does not influence the topological behavior.

We continue to study the effects of In $5s$ orbitals on surface states by calculating the surface bandstructures both with and without In $5s$ orbitals. The surface bandstructures shown in Fig. 3.6 are calculated with the "slab method," where the first-principles TB models of slabs of In- and Sb-substituted Bi_2Se_3 with finite thickness stacked along

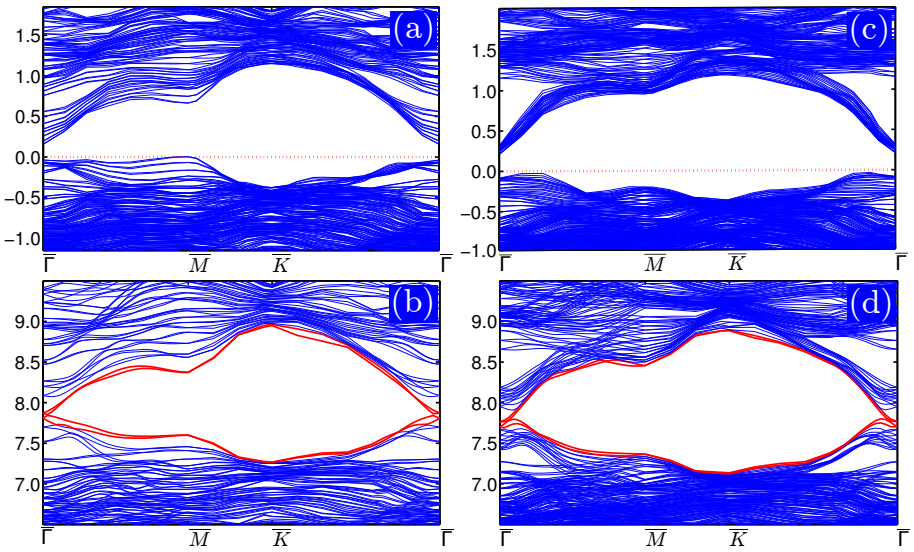

Figure 3.6: Surface bandstructures for $(Bi_{1-x}In_x)_2Se_3$ slabs, plotted in the 2D surface Brillouin zone. (Surface states are shown in red.) (a) 8QL slab for $x = 0.125$. (b) 4QL slab for $x = 0.125$ but with In $5s$ orbitals removed. (c) 12QL slab for energetically favored configuration $C'_{0.25}$ at $x = 0.25$. (d) 8QL slab for $C'_{0.25}$ with In $5s$ orbitals removed. Note split Dirac cones arising at $\bar{\Gamma}$ in (b) and (d).

the [111] direction have been constructed. This is done by extrapolating the matrix elements of the primitive unit cell TB model to multiple QLs along the [111] direction, then truncating at the two surfaces to enforce open boundary conditions. The 2D surface bandstructure is then obtained by directly diagonalizing the TB model of the slab.

It has to be noted that the surface states are not calculated self-consistently by doing such a truncation at the surface, because the Wannier functions close to the surface could be significantly deformed and the hopping parameters between orbitals close to the surface are expected to be different from those deep in the bulk. These effects are not properly included simply by truncating at the slab boundaries. However, we argue that even though these surface effects could be important in determining such details as the exact position of the Dirac point (if present) relative to the bulk CBM, they cannot change the topological character of the surface states, which is what we really care about here.

Table 3.2: Existence of topological surface states vs. impurity concentration x in $(Bi_{1-x}In_x)_2Se_3$ and $(Bi_{1-x}Sb_x)_2Se_3$ based on slab calculations in the Wannier representation.

	0%	6.25%	12.5%	25%	50%	75%	87.5%	100%
With In $5s$	✓	✓	×	×	×	×	×	×
Without In $5s$	✓	✓	✓	✓	×	×	×	×
Sb substitution	✓	-	✓	✓	✓	✓	✓	×

To understand the role of In $5s$ orbitals in the phase-transition process, the surface states are calculated both with and without the In $5s$ orbitals. The results for the configurations with lowest energy (the In-clustered configuration) are summarized in Table 3.2, where '✓' indicates the existence of a Dirac cone around $\overline{\Gamma}$, and × denotes the absence of such a Dirac cone. The thicknesses of the slabs are chosen such that the interference between the states from two opposite surfaces is negligible.

As can be seen in Table 3.2, the In $5s$ orbitals are directly responsible for removing the Dirac cones from the surface spectrum for the $C_{0.125}$ and $C'_{0.25}$ cases. This is shown explicitly in Fig. 3.6, where Dirac cones emerge at $\overline{\Gamma}$ only when the In $5s$ orbitals are removed. (Actually, a close inspection of the figure shows a split pair of Dirac cones contributed by opposite surfaces, where the splitting arises because of broken inversion symmetry due to the pattern of In substitution.) Scanning over the Sb-substituted Bi_2Se_3 supercells from $x = 0.125$ to 0.875, clear signatures of Dirac cones are observed for all of them (not shown here), consistent with the results of the bulk $\mathbb{Z}_2$-index calculations of Fig. 3.3(a).

Based on our supercell calculations, we conclude that this suppression of the topological surface states results from the In $5s$ orbitals, with an In concentration of $x = 0.125$ or $x = 0.25$ being sufficient to remove them entirely.

3.3.6 Disordered spectral functions

The previous superlattice calculations enabled us to capture some important physics of the phase-transition behavior, but it is still difficult to give a precise estimate of the

critical concentrations because of the limited size of the supercells and the approach of studying one particular configuration at a time. Here we use the Wannier representation to construct ensembles representing the disordered systems in much larger supercells, in an attempt to study the effects of disorder in the phase-transition process more realistically and estimate the critical points more accurately.

Two issues arise when disorder is included. First, in a periodic lattice structure without any disorder, the eigenstates of are Bloch states which are perfectly coherent with infinite lifetime, and the wavevector $\mathbf{k}$ is a good quantum number. For such systems with nontrivial band topology, the easiest way to study the topological phase transition is to look at the band structure; a band-gap closure usually implies a phase transition from a topological to trivial insulator. In disordered systems, however, the "band-gap closure" is not so easy to recognize, because the Bloch functions are no longer the eigenstates of the system and a bandstructure is really not well-defined. Secondly, as we know, the $\mathbb{Z}_2$ index is computable for periodic lattices if the information of occupied Bloch states in the entire BZ is given. However, it is a difficult question how to define the $\mathbb{Z}_2$ index and determine the topological behavior of a realistic disordered system.

Our answer to the first issue is to look at the disorder-averaged spectral functions computed from a large supercell with different impurity configurations, but unfolded back to the BZ of the primitive unit cell. If the disorder is weak, one should see a sharp spectrum with narrow lifetime broadening, which means the quantum states would remain coherent over long distances. For strongly disordered systems, however, it is expected that the spectral functions should be strongly smeared out due to the strong randomness of the impurity scattering, and the quantum states would be localized around the impurities with a relatively short localization length.

We propose the "$\mathbb{Z}_2$-index statistics" to address the second problem. To be specific, several different impurity configurations are generated in our calculations at each impurity composition x, forming a representative ensemble of the disordered system, and

we then compute the strong $\mathbb{Z}_2$ index for each configuration. If each configuration in the disorder ensemble is equally weighted, then when over half of the configurations are in the $\mathbb{Z}_2$-odd phase, we say that the disordered system can be statistically considered as a TI.

To calculate the disordered spectral functions, we use the Wannier effective-Hamiltonian approach as well as the technique of unfolding first-principle supercell bandstructures (spectral functions) as proposed by Ku and coworkers [115, 116].We note in passing that a similar technique has been proposed by Popescu and Zunger [131]. To be explicit, we first construct a TB model for a 4×4×3 supercell of pure Bi_2Se_3 whose matrix elements are extrapolated from the primitive-cell Bi_2Se_3 TB model. The 4×4×3 supercell of the bulk material acts as the reference system in which the Bi atoms are randomly substituted by the impurity atoms. For a 4×4×3 supercell, there are 240 atoms of which 96 are Bi-like, so the impurity composition x can be varied on the scale of one percent, enabling us to determine the critical point x_c with high precision.

The next procedure is to extract the Hamiltonian of a single impurity defined under the same WF basis. This is done by working in a small supercell (2×2×1 in our case) and subtracting the pure bulk-material Hamiltonian H^0 from the Hamiltonian H^s of a supercell containing one substituted impurity of type s. To set the notation, we label (5-atom) cells within the supercell by l, sites within the cell as τ, and orbitals within the cell as m. Then the impurity potential is constructed as

$$\Delta^s_{lm,l'm'}(\tau_s) = (H^s_{lm,l'm'}(0\tau_s) - H^0_{lm,l'm'})\, P_{lm,l'm'}(\tau_s)\,. \tag{3.1}$$

This describes the change in the on-site energy if $(lm) = (l'm')$, or in the hopping if $(lm) \neq (l'm')$, induced by the presence of the impurity of type s on site τ_s in the central cell $l_s = 0$ of the small supercell. We define a "partition function" [116] $P_{lm,l'm'}(\tau_s)$ that is used to partition the contribution of the single impurity from the super-images in the neighboring supercells, such that the single impurity Hamiltonian is not influenced by the artificial periodicity of the supercell. In our calculations this partition function

is chosen as

$$P(d) = \begin{cases} e^{-(d/r_0)^8} & \text{if } d \le d_{\rm c} \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

where $d = d_{lm,l'm'}(\tau_s) \equiv |\mathbf{r}_{lm} - \mathbf{r}_{0\tau_s}| + |\mathbf{r}_{l'm'} - \mathbf{r}_{0\tau_s}|$ is chosen as a measure of the "distance" from the hopping matrix element $(lm, l'm')$ to the impurity site located at $\mathbf{r}_{0\tau_s}$ [116]. Here we choose $d_{\rm c} = 8.69\,$Å and $r_0 = 7.86\,$Å. (We find that if $d_{\rm c} > 8.5\,$Å and r_0 is chosen slightly smaller than or equal to $d_{\rm c}$, the impurity Hamiltonian becomes insensitive to small variations of $d_{\rm c}$ and r_0.) Our partition scheme has been tested to be able to reproduce the first-principles $2\times2\times1$ supercell bandstructures at $x = 0.125$ and 0.25.

We extract this impurity potential once and for all for an In atom substituting for the top Bi atom in the quintuple-layer ($s = 1$) and again when it substitutes for the bottom Bi atom ($s = 2$). Then for a particular impurity configuration $\mathcal{R} = \{l_1 s_1, l_2 s_2, \ldots\}$ of the $4 \times 4 \times 3$ supercell, where $l_j s_j$ specifies the subcell and type of impurity and j runs over the impurities in the supercell, the effective Hamiltonian is taken as a linear superposition of the reference-system Hamiltonian H^0 and the single-impurity Hamiltonians residing on the specified sites in the large supercell, i.e.,

$$H^{\mathcal{R}}_{lm,l'm'} = H^0_{lm,l'm'} + \sum_j \Delta^{s_j}_{(l-l_j)m,(l'-l_j)m'}(\tau_j)\,. \tag{3.3}$$

The linear superposition of the matrix elements of different TB Hamiltonians is well-defined only when these Hamiltonians are treated under the same WF basis. In other words, each of the orbitals from the large supercell with impurities should map appropriately to the corresponding orbitals of the unperturbed reference system. For this reason, we skip the maximal localization procedure when generating the WFs, and instead simply use the projection method to generate a basis that remains in close correspondence to the atomic-like orbitals. Once the effective Hamiltonians has been obtained for an ensemble of impurity configurations representing a given concentration

x, we calculate the spectral function for each, and unfold it from the highly compressed supercell BZ into the primitive-cell BZ [115]. Finally, the ensemble average of the unfolded spectral functions can then be taken to reflect the effects of disorder on the original bulk electronic states.

To be specific, let $A_N(\omega, \mathbf{K})$ be the spectral function at energy ω associated with band N in the supercell, with $\mathbf{K}$ specifying the wavevector in the small supercell BZ, given by the imaginary part of the retarded Green's function operator G via $A = -\pi^{-1} \operatorname{Im} G = -\pi^{-1} \operatorname{Im}(\omega + i\eta - H)^{-1}$, where H is the supercell Hamiltonian and $\eta > 0$ is a small artificial smearing factor. Then, to unfold the supercell spectral function onto a complete set of primitive-cell Bloch states, one can expand the primitive-cell spectral function $A_n(\omega, \mathbf{k})$ in terms of the supercell spectral functions as

$$A_n(\omega, \mathbf{k}) \;=\; \sum_{N\mathbf{K}} |\langle \psi_{N\mathbf{K}} | \psi_{n\mathbf{k}} \rangle|^2 A_N(\omega, \mathbf{K}) \,, \tag{3.4}$$

where $|\psi_{n\mathbf{k}}\rangle$ and $|\psi_{N\mathbf{K}}\rangle$ are the primitive-cell and supercell Bloch states respectively, and n and $\mathbf{k}$ represent the band index and wavevector of the primitive cell. One can solve for the coefficient $\langle \psi_{N\mathbf{K}} | \psi_{n\mathbf{k}} \rangle$ within the Wannier basis provided that the supercell Hamiltonian is defined under a set of WFs having a clear one-to-one mapping with the primitive-cell WFs by primitive-cell lattice translations, as can be realized by using simple projection for the Wannier construction [2].

In our calculations, 16 configurations are generated for $(Bi_{1-x}In_x)_2Se_3$ at each impurity composition, whereas eight configurations are generated for each x of $(Bi_{1-x}Sb_x)_2Se_3$ due to the much weaker effect of disorder. For $(Bi_{1-x}Sb_x)_2Se_3$ the configurations are generated randomly, as different configurations seem to be equally favored energetically. For $(Bi_{1-x}In_x)_2Se_3$, however, the configurations are generated using the Metropolis Monte Carlo method according to a proper Boltzmann weight in order to reflect the

[2] Expressed under Wannier basis, the summation over $\mathbf{K}$ in Eq. (3.4) becomes unnecessary, because there is a Kronecker-δ relationship between $\mathbf{k}$ and $\mathbf{K}$ in the expression for $\langle \psi_N(\mathbf{K}) | \psi_n(\mathbf{k}) \rangle$, thus removing the summation.

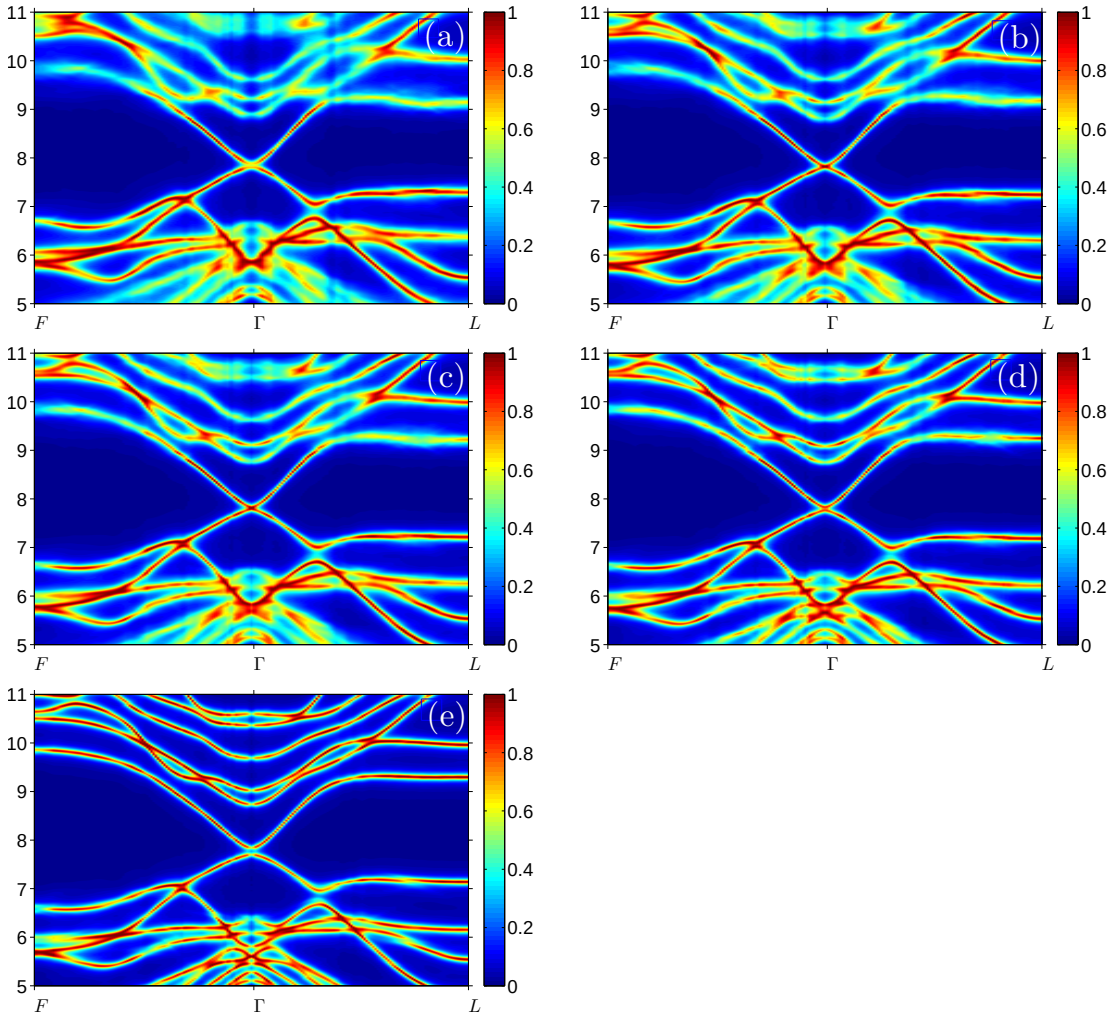

Figure 3.7: Disordered spectral functions of $(Bi_{1-x}Sb_x)_2Se_3$ unfolded into the primitive-cell BZ. (a) $x=68\%$. (b) $x=78\%$. (c) $x=83\%$. (d) $x=88.5\%$. (e) $x=99\%$.

tendency of In segregation. The Boltzmann weight is proportional to $e^{-(E_pN_p)/(k_BT_g)}$, where $E_p = [E(C'_{0.25}) - E(C_{0.25})]/2$ is defined as the "paring energy" of In atoms, N_p is the number of In pairs in a particular configuration, k_B is the Boltzmann constant, and $T_g = 850°C$ is taken as the growth temperature of the In-substituted Bi_2Se_3 sample.

The disordered spectral functions of $(Bi_{1-x}Sb_x)_2Se_3$ are shown in Fig. 3.7, with an artificial Lorentzian broadening of 2 meV. At $x = 68\%$, the spectral gap is visible, but already very small, suggesting that the system is approaching the critical point. As x is increased to 78%, a sharp Dirac cone is observed, which remains robust from 78% to 83%. At $x = 88.5\%$, the band gap reopens, meaning that the system is in the

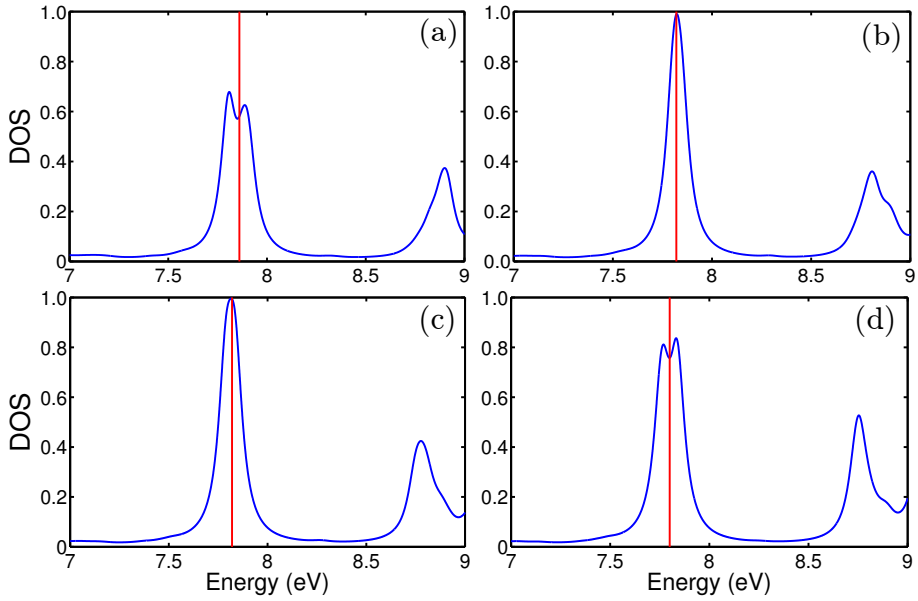

Figure 3.8: Disordered spectral functions of $(Bi_{1-x}Sb_x)_2Se_3$ at the Γ point in the primitive-cell BZ for (a) $x = 68\%$, (b) $x = 78\%$, (c) $x = 83\%$, and (d) $x = 88.5\%$. Vertical (red) line indicates Fermi energy. Distinct VBM and CBM peaks are still visible in the topological phase in (a), merge in (b) and (c), and reappear in (d) as the gap reopens in the normal phase.

NI phase, and this topologically trivial phase becomes more robust as x goes to 99% with a more visible gap. One may notice that the effect of disorder is weak during the phase-transition process, and the semi-metallic behavior at criticality is rather sharp.

The spectral functions of $(Bi_{1-x}Sb_x)_2Se_3$ at the Γ point (in the primitive-cell BZ) are plotted in Fig. 3.8. At $x = 68\%$ there are two peaks around the Fermi level, indicating that the CBM and VBM are still separated, and the critical point has not been reached yet. At $x = 78\%$ and 83%, the two peaks from the conduction and valence bands merge into one, suggesting the system becomes a semimetal. As x goes to 88.5%, the gap opens up again. From these results it appears that there is a kind of "critical plateau" for x between $\sim$78% and $\sim$83%.

This critical behavior observed for $(Bi_{1-x}Sb_x)_2Se_3$ is at variance with the general expectation for the topological phase-transition behavior in the Bi_2Se_3 class of TIs, where the system becomes a critical Dirac semimetal at one point in the parameter space (here it is the impurity composition x) and then becomes insulating again immediately

after the gap closure. This deserves discussion.

We cannot exclude the possibility that numeric uncertainties play a role here. In $(Bi_{1-x}Sb_x)_2Se_3$, the band gap varies quite slowly as a function of x, with a change of impurity composition of 5% corresponding to a band gap change of only $\sim$0.03 eV. Thus, the critical point could be hidden by disorder and artificial smearing, such that the system looks metallic even if a very small band gap has opened up. Moreover, finite-size effects may be important. In a disordered system we expect that the localization length evaluated in the middle of the mobility gap should grow as the mobility gap shrinks and the system approaches the critical metallic state. As one approaches the critical point at which this mobility gap vanishes, the localization length may exceed the size of the supercell and the states in neighboring supercells may overlap and behave like extended states.

However, it is also possible that a finite window of metallic phase is physically correct. Since the topologically nontrivial and trivial insulating configurations compete with similar weight near criticality, the system may remain in the metallic phase until one of the two insulating phases comes to dominate. Support for this picture can be drawn from Ref. [132], in which careful numerical simulations on a disordered lattice model showed a finite-width region of metallic phase as the system was driven from the TI to the NI phase with increasing disorder strength while other parameters were held fixed. In our case the disorder strength remains approximately constant, but the ratio of disorder strength to energy gap varies with x, so that a metallic plateau may still be expected. We leave these questions as avenues to pursue in future research.

The disorder-averaged spectral functions for $(Bi_{1-x}In_x)_2Se_3$ with the same artificial broadening are plotted in Fig. 3.9. At 8.3%, the band-inversion character is still obvious (note the tilde-like shape of the highest occupied bands around Γ) with the spectral gap unclosed, which implies the system may still stay in a TI phase. At 12.5% the spectral gap becomes hard to recognize. When it comes to 14.6%, 16.7% and 18.8%,

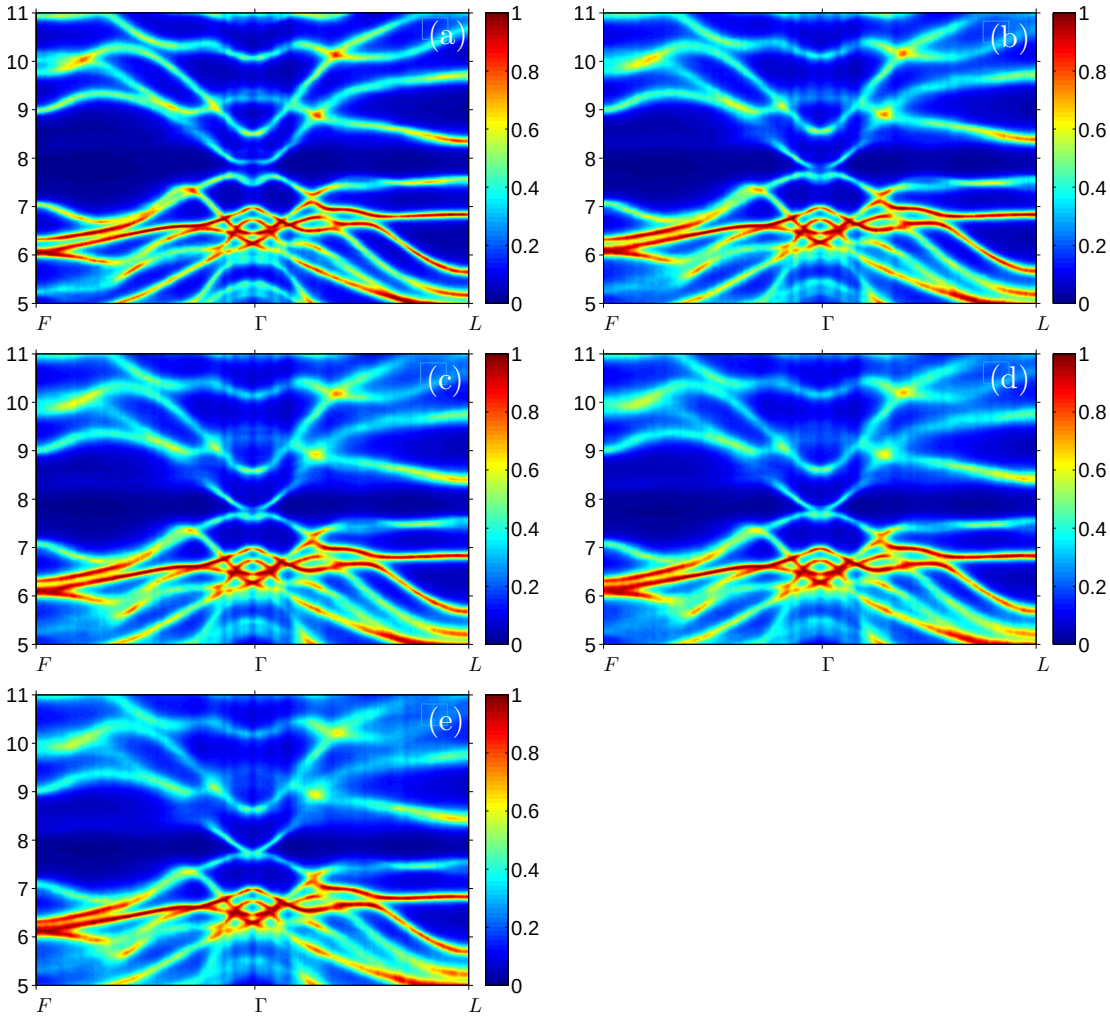

Figure 3.9: Disordered spectral functions of $(Bi_{1-x}In_x)_2Se_3$ unfolded into the primitive-cell BZ. (a) $x=8.3\%$. (b) $x=12.5\%$. (c) $x=14.6\%$. (d) $x=16.7\%$. (e) $x=18.8\%$.

the spectral gap is almost completely unrecognizable, which means the system is pretty close to the critical point.

Moreover, in sharp contrast with the behavior of $(Bi_{1-x}Sb_x)_2Se_3$, the effect of disorder in $(Bi_{1-x}In_x)_2Se_3$ is very strong. It can be seen from Fig. 3.9 that the original energy bands are strongly smeared out. Different Bloch states are mixed together, as would be expected if localized eigenstates are formed centered on the substituted In atoms. It is difficult to identify the critical point simply by looking at the disordered spectral functions, because the Dirac semi-metallic behavior is not as obvious as in $(Bi_{1-x}Sb_x)_2Se_3$. Therefore, we calculate the $\mathbb{Z}_2$ index of each configuration from

$x = 8\%$ to 18.8%. By inspecting the statistical behavior of the resulting $\mathbb{Z}_2$ indices, as described next, we conclude that the critical point of $(Bi_{1-x}In_x)_2Se_3$ is around 17%.

Our theoretical prediction of x_c for $(Bi_{1-x}In_x)_2Se_3$ is somewhat higher than the experimental values, estimated to be 3%-7% according to Brahlek *et al.* [41] and ∼6% according to Wu *et al.* [40] We attribute our overestimate of x_c to both the use of standard DFT methods and the absence of impurity-impurity correlation terms in Eq. (3.3). Regarding the former, we would expect to get an x_c more consistent with the experimental results if hybrid functionals or more advanced many-body first-principles methods were used in the calculations, which unfortunately becomes expensive for large supercells. Regarding the latter, we expect that the In clustering effects would be treated more accurately if we would go beyond a simple superposition of one-body impurity potentials and include many-body terms in the impurity cluster expansion when constructing the effective Hamiltonian. However, this too would carry a large computational cost due to the anisotropic nature of the two-body impurity-impurity interactions and the fact that higher-body terms may also be important.

3.3.7 $\mathbb{Z}_2$-index statistics

The $\mathbb{Z}_2$ indices of a 3D band insulator are well defined for a perfect periodic lattice with time-reversal symmetry. For disordered systems, however, the topological indices are much harder to calculate. A promising approach is the use of non-commutative algebra [133, 134, 135, 136], but to date this has generally been applied to simple models, and its applicability to realistic disordered materials has not been demonstrated.

Here we attempt to determine the topological indices of a disordered time-reversal invariant insulating system using a more straightforward approach: we calculate the strong $\mathbb{Z}_2$ index (with periodic boundary conditions on the supercell) for each impurity configuration in the statistical ensemble, thus determining the topological properties of the disordered system from a statistical point of view. As long as the configurations

Table 3.3: $\mathbb{Z}_2$ statistics of $(Bi_{1-x}In_x)_2Se_3$. The entries in the first, second, and third rows indicate respectively the number of $\mathbb{Z}_2$-odd, $\mathbb{Z}_2$-even, and metallic configurations drawn from a 16-member ensemble.

	8.3%	12.5%	14.6%	16.7%	18.8%
$\mathbb{Z}_2$ odd	15	16	14	11	5
$\mathbb{Z}_2$ even	0	0	0	3	9
Metallic	1	0	2	2	2

are sampled in such a way that each contributes equally to the statistical ensemble, then we define the system as $\mathbb{Z}_2$-odd (i.e., a strong TI) if over half of the configurations are $\mathbb{Z}_2$-odd, and normal otherwise. As mentioned in Subsec. 3.3.6, the impurity configurations of $(Bi_{1-x}In_x)_2Se_3$ are generated using the Metropolis Monte Carlo method based on a Boltzmann weight defined by the In-clustering energy. As a result, the tendency of In segregation is reflected in the number of generated distributed vs. clustered configurations, rather than by manual assignment of weights. Thus we consider each configuration to be equally weighted, satisfying the criterion stated above.

The strong $\mathbb{Z}_2$ index statistics of several $(Bi_{1-x}In_x)_2Se_3$ configurations are shown in Table 3.3. For x between 16.7 and 18.8%, the number of $\mathbb{Z}_2$-odd configurations drops from eleven to five, so we estimate x_c to be approximately 17%.

3.4 summary and outlook

To summarize, we have studied the topological phase transitions in $(Bi_{1-x}In_x)_2Se_3$ and $(Bi_{1-x}Sb_x)_2Se_3$ using two approaches, the direct application of first-principles calculations on small supercells, and a Wannier-based modeling approach that allows for a more realistic treatment of disorder in large supercells. Based on the former approach, the x_c of $(Bi_{1-x}In_x)_2Se_3$ is slightly less than 12.5%, while that of $(Bi_{1-x}Sb_x)_2Se_3$ is even above 87.5%. A VCA treatment of $(Bi_{1-x}Sb_x)_2Se_3$ predicts $x_c \sim 65\%$; this is not in perfect agreement with the prediction from the supercell calculations, but both of

them are much higher than that of $(Bi_{1-x}In_x)_2Se_3$. From the results of realistic disordered calculations, we found that x_c is ∼17% for $(Bi_{1-x}In_x)_2Se_3$, while it is ∼78-83% for $(Bi_{1-x}Sb_x)_2Se_3$. The critical concentrations are determined from disorder-averaged spectral functions and $\mathbb{Z}_2$-index statistics. It is concluded that in $(Bi_{1-x}Sb_x)_2Se_3$, the band gap at Γ decreases almost linearly with increasing x, corresponding to the reduction in average SOC strength, with only a very weak disorder effect. For $(Bi_{1-x}In_x)_2Se_3$, on the other hand, the In $5s$ orbitals tend strongly to suppress the topological band inversion even at very low impurity concentrations, so that the phase transition is drastically accelerated as a function of increasing x.

In the case of $(Bi_{1-x}Sb_x)_2Se_3$, we observed a critical plateau from $x \sim 78\%$ to $x \sim 83\%$. As discussed in Subsec. 3.3.6, it is difficult to say whether this intermediate metallic phase is just an artifact of numerical limitations such as finite-size effects, or is a true feature of the physics. Further theoretical and experimental work is needed to clarify what happens in this critical region.

We also find a tendency of the In (but not Sb) atoms to segregate. This In clustering effect could help clarify some aspects of the topological phase transition in $(Bi_{1-x}In_x)_2Se_3$, as for example by suggesting a scenario in which the phase transition may happen locally, instead of homogeneously as in the usual linear gap-closure picture. One can imagine that as In atoms are implanted into bulk Bi_2Se_3, isolated In clusters would start to emerge, inside which the system is topologically trivial. As more and more Bi atoms are substituted by In, these isolated In "islands" become connected to each other, and the topological phase transition happens when the percolation threshold is reached.

Our results for $(Bi_{1-x}In_x)_2Se_3$ provide a physical explanation for the observed low transition concentration in several recent experiments on $(Bi_{1-x}In_x)_2Se_3$[40, 41], and the results on $(Bi_{1-x}Sb_x)_2Se_3$ may give predictions for future experimental works.

The techniques used in this chapter provide a powerful methodology that may be

used to carry out theoretical explorations of other types of disordered topological systems. For example, interesting physics is anticipated in a TI whose bandstructure is mostly contributed by p orbitals while substituting with impurities having d or f orbitals. It could also be interesting to investigate the effects of magnetic impurities, not only on the surface states and their spin textures [137, 138], but on the bulk topological transition as well. We thus hope that these methods will enable the search for new materials and systems with non-trivial topological properties in strongly disordered alloy systems.

This work is supported by NSF Grant DMR-10-05838. We are grateful to M. Taherinejad and K. Garrity for useful discussions, I. Souza for important technical assistance with the codes, and W. Wu and his collaborators for sharing their unpublished experimental data.

Chapter 4

Weyl semimetals from noncentrosymmetric topological insulators

In this chapter, we study the problem of phase transitions from 3D topological to normal insulators without inversion symmetry. In contrast with the conclusions of some previous work, we show that a Weyl semimetal always exists as an intermediate phase regardless of any constriant from lattice symmetries, although the interval of the critical region is sensitive to the choice of path in the parameter space and can be very narrow. We demonstrate this behavior by carrying out first-principles calculations on the noncentrosymmetric topological insulators $LaBiTe_3$ and $LuBiTe_3$ and the trivial insulator BiTeI. We find that a robust Weyl-semimetal phase exists in the solid solutions $LaBi_{1-x}Sb_xTe_3$ and $LuBi_{1-x}Sb_xTe_3$ for $x \approx 38.5 - 41.9\%$ and $x \approx 40.5 - 45.1\%$ respectively. A low-energy effective model is also constructed to describe the critical behavior in these two materials. In BiTeI, a Weyl semimetal also appears with applied pressure, but only within a very small pressure range, which may explain why it has not been experimentally observed.

4.1 Introduction

The significance of topology in determining electronic properties has became widely appreciated with the discovery of the integer quantum Hall effect and been highlighted further by the recent interest in topological insulators (TIs) [21, 22, 33, 34]. In topological band theory, a topological index, such as the Chern number or the $\mathbb{Z}_2$ index, is well-defined only for gapped systems, and the topological character is signaled by

the presence of novel gapless surface states which cannot exist in any isolated 2D system [33, 34]. Recently, the concept of topological phases is further generalized to 3D bulk gapless systems, whose topological behavior is protected by lattice translational symmetry, known as the Weyl semimetal (WSM) [139, 140, 43, 45, 141].

A Weyl semimetal is characterized by a Fermi energy that intersects the bulk bands only at one or more pairs of band-touching points (BTPs) between nondegenerate valence and conduction bands. This can occur in the presence of spin-orbit coupling (SOC), typically in a crystal with broken time-reversal or inversion symmetry but not both, so that the pairs are of the form ($\mathbf{k}_0$, $-\mathbf{k}_0$) in the Brillouin zone (BZ). The effective Hamiltonian around a single BTP $\mathbf{k}_0$ can be written as $H(\mathbf{k}) = f_0(\mathbf{k}) + \mathbf{f}(\mathbf{k}) \cdot \boldsymbol{\sigma}$, where f_0 and $\mathbf{f}$ are scalar and vector functions respectively of wavevector in the BZ and the σ_j are the Pauli matrices acting in the two-band space. If one expands the coefficient $\mathbf{f}(\mathbf{k})$ to linear order around $\mathbf{k}_0$, one gets a Hamiltonian having the form of the Weyl Hamiltonian in relativistic quantum mechanics after a coordinate transformation in $\mathbf{k}$ space. If the sign of the determinant of the Jacobian that describes the coordinate transformation is positive (negative), we call the BTP as a Weyl node with positive (negative) chirality, and the low-energy excitations around such a Weyl node provide a condensed-matter realization of left-handed (right-handed) Weyl fermions.

These pairs of Weyl nodes are topologically protected in the sense that they are robust against small perturbations, which can be see from the codimension argument as follows. One can introduce a parameter λ that acts as a perturbation on the BTP, and let both f_0 and $\mathbf{f}$ to be dependent on λ. In order to get a band touching at $(\mathbf{k}_0, \lambda_0)$, the three coefficients $\mathbf{f} \equiv (f_x, f_y, f_z)$ have to vanish. However, since there are four degrees of freedom, if $\lambda_0 \to \lambda_0 + \delta\lambda$, instead of opening a gap, the Weyl node would just shift slightly in momentum to compensate for the perturbation. In fact, there is no way to remove a Weyl node unless two Weyl nodes with opposite chirality annihilate each other.

If the two Weyl nodes are aligned in energy due to either time-reversal or some lattice symmetry, and the bands are filled right up to the Weyl nodes, then the Fermi energy would be locked there regardless of weak perturbations. That is, the Fermi level could be slightly shifted upward (downward) due to some weak perturbation, such that there is an electron-like (hole-like) Fermi surface, then there must also be a hole-like (electron-like) Fermi surface to conserve the total number of electrons, which is impossible in such a a semimetal. It follows that the low-energy physics in the Weyl semimetal is completely dominated by the linearly dispersing states around the Weyl nodes, which leads to interesting surface states and transport properties.

The presence of Weyl nodes in the bulk bandstructure is responsible for the presence of Fermi arcs at the surface, which can be understood as follows [43, 45], Consider a small loop in the 2D surface BZ that encloses the projection along k_z of one Weyl point. When translated along k_z, this loop traces out a surface in the 3D BZ, and the application of Gauss's theorem implies that the Chern number on this surface must equal the chirality of the enclosed Weyl node. It follows that as (k_x, k_y) is carried around the loop, a single electron is pumped up to (or down from) the top surface, and this is only consistent with charge conservation if a single surface state crosses the Fermi energy E_{F} during the cycle. Since this argument applies for an arbitrary loop, surface states must exist at E_{F} along some arc emerging from the surface-projected Weyl point. If there is another Weyl node with opposite chirality, then the Chern number can vanish once the cylinder encloses both of the nodes, such that the Fermi arc would only extend between the two projected Weyl nodes [43, 45]. It is also interesting to note that the surface states resided on top and bottom surfaces are linearly dispersed in opposite directions, and they remain robust even when the Fermi energy is away from the bulk Weyl nodes [142].

In a WSM with broken time-reversal (TR) symmetry, there is also a non-zero anomalous Hall conductivity (AHC) that is closely related to the positions of the Weyl nodes

[143, 44, 144]. Consider the AHC σ_{xy} with the crystal oriented such that the third primitive reciprocal vector is along z. If we track the Chern number C_z of a 2D slice normal to z in the 3D BZ, we must find that it changes by ± 1 whenever k_z passes a Weyl node, with the sign depending on the chirality of the node. In the simplest case, if $C_z = 1$ for k_z between the two Weyl nodes and zero elsewhere, then the AHC is just proportional to the separation of the two Weyl nodes in k_z. This is also interpreted as a consequence of the "chiral anomaly" in a WSM [145]. Other interesting transport phenomena can arise due to the chiral anomaly. For example, if a magnetic field $\mathbf{B}$ is applied to a WSM in the z direction, Landau levels will be formed in the (x, y) plane. The zeroth Landau level disperses linearly along k_z, but in opposite directions for Weyl nodes with opposite chirality. As a result, if an electric field $\mathbf{E}$ is applied along z, electrons would be pumped from one Weyl node to the other at a rate proportional to $\mathbf{E}\cdot\mathbf{B}$, with the Fermi arcs serving as a conduit [139, 146, 45, 144].

As discussed above, a WSM requires the breaking of either TR or inversion symmetry. Many of the previous works are focused on WSMs without TR symmetry, such as in pyrochlore iridates [43], magnetically doped TI multilayers [44], and $Hg_{1-x-y}Cd_xMn_yTe$ [147]. In this chapter, we study the WSM with preserved TR symmetry but broken inversion symmetry.

It was argued some time ago that the $\mathbb{Z}_2$-odd and $\mathbb{Z}_2$-even phases of a noncentrosymmetric insulator should always be bridged by a critical WSM phase [47, 46]. If the transition is described by some adiabatic parameter λ, then as λ increases one expects first the appearance of m higher-order BTPs in the half BZ (and another m at the time-reversed points), where $m = 1$ is typical of low-symmetry systems while $m > 1$ can occur when, e.g., rotational symmetries are present. These higher-order BTPs generally have quadratic dispersion in one direction while remaining linear in the other two, and are non-chiral; we refer to such a point henceforth as a "quadratic BTP." As λ increases, each quadratic BTP splits to form a pair of Weyl nodes ($4m$

altogether), which then migrate through the BZ and eventually annihilate at a second critical value of λ after exchanging partners. The previous work demonstrated that this process inverts the strong $\mathbb{Z}_2$ index if m is odd [47, 46]. Recently, however, Yang *et al.* claimed that for systems with certain high-symmetry lines in the BZ, the phase transition could occur at a unique critical value of λ at which the bands would touch and immediately reopen, instead of over some finite interval in λ, even when inversion symmetry is absent [148]. These authors suggested that BiTeI under pressure could serve as an example to support their claim [149, 148].

In this chapter, we address this issue carefully. We show that an intermediate critical WSM phase should always exist for any topological phase transition (TPT) between a normal and a $\mathbb{Z}_2$-odd insulating phase. We find however that the width of the critical WSM phase can be sensitive to the choice of path in parameter space and can sometimes be very small. To justify our conclusions, we take specific materials as examples. We first study the TPT in the solid solutions $LaBi_{1-x}Sb_xTe_3$ and $LuBi_{1-x}Sb_xTe_3$ using the virtual crystal approximation, where the phase transition is driven by Sb substitution. The parent compounds at $x = 100\%$, $LaBiTe_3$ and $LuBiTe_3$, are hypothetical noncentrosymmetric materials that are predicted to be strong topological insulators in Ref. [150] and in the present work respectively. Instead, the end members $LaSbTe_3$ and $LuSbTe_3$ at $x = 0\%$ are trivial insulators [150]. We find that a WSM phase is obtained when x is in the range of about 38.5-41.9% for $LaBiTe_3$ and 40.5-45.1% for $LuBiTe_3$. We further construct a low-energy effective model to describe the topological and phase-transitional behavior in this class of materials. We also revisit the TPT of BiTeI driven by applied pressure, where a WSM phase has not previously been observed [149, 151]. Based on our calculations, we find that a small interval of WSM phase does actually intervene as increasing pressure drives the system from the trivial to the topological phase.

This chapter is organized as follows. In Sec. 4.2 we derive the general behavior of

TPTs in noncentrosymmetric insulators and point out some deficiencies in the discussion of BiTeI by Yang *et al* [148]. In Sec. 4.3 we describe the lattice structures and basic topological properties of the materials, as well as the numerical methods used in the realistic-material calculations, especially the methods used in modeling the alloyed and pressurized systems and in searching for BTPs in the BZ. In Sec. 4.4 we present the results for $LaBi_{1-x}Sb_xTe_3$, $LuBi_{1-x}Sb_xTe_3$ and BiTeI, and discuss the sensitivity to the choice of path. In Sec. 4.5, we summarize our work.

4.2 Topological transition in noncentrosymmetric insulators

4.2.1 General behavior

We consider the problem of TPTs in noncentrosymmetric insulators in the most general case. In the space of the two bands which touch at the TPT, the system can be described by the effective Hamiltonian

$$H(\mathbf{k},\lambda) = f_x(\mathbf{k},\lambda)\sigma_x + f_y(\mathbf{k},\lambda)\sigma_y + f_z(\mathbf{k},\lambda)\sigma_z\,, \tag{4.1}$$

where λ is the parameter that drives the TPT and $\sigma_{x,y,z}$ are the three Pauli matrices defined in the space spanned by the highest occupied and the lowest unoccupied states at $\mathbf{k}$. Since we study the TPT between two insulating phases, we can assume without loss of generality that the system is gapped for $\lambda<\lambda_0$, and that the first touching that occurs at $\lambda=\lambda_0$ takes place at $\mathbf{k}=\mathbf{k}_0$. In other words, $f_i(\mathbf{k}_0,\lambda_0)=0$, $i=x,y,z$. Then we ask what happens if $\mathbf{k}_0\to\mathbf{k}_0+\mathbf{q}$ and $\lambda_0\to\lambda_0+\delta\lambda$.

We first expand the coefficients $\mathbf{f}$ around $(\mathbf{k}_0,\lambda_0)$ as $\mathbf{f}=\mathbf{J}\cdot\mathbf{q}+\mathbf{\Lambda}\,\delta\lambda$, where $\mathbf{q}=\mathbf{k}-\mathbf{k}_0$, $\delta\lambda=\lambda-\lambda_0$, $\mathbf{J}$ is the Jacobian with matrix elements $J_{ij}=(\partial f_i/\partial k_j)|_{\mathbf{k}_0,\lambda_0}$, and $\mathbf{\Lambda}$ is a 3-vector with components $\Lambda_i=(\partial f_i/\partial\lambda)|_{\mathbf{k}_0,\lambda_0}$. A natural set of momentum-space coordinates can be defined in terms of the eigensystem $\mathbf{J}\cdot\mathbf{v}_i=J_i\mathbf{v}_i$. Defining

$\mathbf{q} = \sum_i p_i \, \mathbf{v}_i$ and $\mathbf{u}_i = J_i \mathbf{v}_i$, we obtain

$$\mathbf{f} = \sum_{i=1}^{3} p_i \, \mathbf{u}_i + \delta\lambda \, \mathbf{\Lambda}. \tag{4.2}$$

Following the argument of Yang *et al* [148], the Jacobian matrix $\mathbf{J}$ has to be singular at $(\mathbf{k}_0, \lambda_0)$ because otherwise there would be band touching even when $\lambda < \lambda_0$, contradicting the assumption that the system is insulating for $\lambda < \lambda_0$. This implies that at least one of the eigenvalues of $\mathbf{J}$ is zero. We assume for the moment that the others are non-zero, i.e., that $\mathbf{J}$ is rank-2, and let it be the first eigenvalue that vanishes. Since the p_1-dependence of $\mathbf{f}$ then vanishes at linear order, we again follow Ref. [148] by including a second-order term to obtain

$$\mathbf{f} = p_2 \, \mathbf{u}_2 + p_3 \, \mathbf{u}_3 + \delta\lambda \, \mathbf{\Lambda} + p_1^2 \, \mathbf{w} \tag{4.3}$$

where $\mathbf{w} = (1/2)\partial^2 \mathbf{f} / \partial^2 p_1 |_{\mathbf{k}_0, \lambda_0}$. Now we also have the freedom to carry out an arbitrary rotation in the pseudospin representation of the two-band space. That is, we redefine f_i to be the component in the pseudospin in direction $\mathbf{e}_i$, with $\mathbf{e}_3$ given by $(\mathbf{u}_2 \times \mathbf{u}_3)/|\mathbf{u}_2 \times \mathbf{u}_3|$, and $\mathbf{e}_1$ and $\mathbf{e}_2$ chosen to form an orthonormal frame with $\mathbf{e}_3$. Then u_{23} and u_{33} vanish, and we can write explicitly that

$$\begin{aligned} f_1 &= p_2 \, u_{21} + p_3 \, u_{31} + \delta\lambda \, \Lambda_1 + p_1^2 \, w_1, \\ f_2 &= p_2 \, u_{22} + p_3 \, u_{32} + \delta\lambda \, \Lambda_2 + p_1^2 \, w_2, \\ f_3 &= \delta\lambda \, \Lambda_3 + p_1^2 \, w_3. \end{aligned} \tag{4.4}$$

We assume $\Lambda_3 / w_3 < 0$, since otherwise there are solutions at negative λ. Then at positive λ, there are always two solutions $p_1 = \pm\sqrt{-\delta\lambda\Lambda_3 / w_3}$ at which $f_3 = 0$. Plugging this into the expressions for f_1 and f_2 in Eq. (4.4), we can obtain p_2 and p_3 by solving the linear system

$$\begin{bmatrix} u_{21} & u_{31} \\ u_{22} & u_{32} \end{bmatrix} \begin{bmatrix} p_2 \\ p_3 \end{bmatrix} + \begin{bmatrix} \Lambda_1 - w_1 \Lambda_3 / w_3 \\ \Lambda_2 - w_2 \Lambda_3 / w_3 \end{bmatrix} \delta\lambda = 0. \tag{4.5}$$

Solutions of the above equation always exist as long as the Jacobian matrix $\mathbf{J}$ is of rank two, which means a critical WSM should always exist in the absence of a special lattice symmetry that would lower the rank of $\mathbf{J}$. From the above it also follows that at the critical $\lambda = \lambda_0$ the dispersion around $\mathbf{k}_0$ is quadratic in p_1 and linear in p_2 and p_3, and that for larger λ the Weyl point displacements scale like $|p_1| \sim \sqrt{\delta\lambda}$ and $|p_{2,3}| \sim \delta\lambda$. The same conclusions in the rank-two case have been obtained by Murakami *et al.* [47] and restated by Yang *et al.* [148].

If the Jacobian matrix $\mathbf{J}$ turns out to be rank-one instead at λ_0, then the bands would first close at a doubly-quadratic BTP. That is, there would be two vanishing eigenvalues of the Jacobian matrix (which we take to be the first and second), and the dispersion would be quadratic in p_1 and p_2 and linear in p_3. This implies that only the second-order terms associated with p_1 and p_2 need to be included in Eq. (4.3), yielding

$$\begin{aligned} f_1 &= p_3\, u_{31} + \delta\lambda\, \Lambda_1 + p_1^2\, w_1^{11} + p_2^2\, w_1^{22} + 2p_1 p_2\, w_1^{12}, \\ f_2 &= p_3\, u_{32} + \delta\lambda\, \Lambda_2 + p_1^2\, w_2^{11} + p_2^2\, w_2^{22} + 2p_1 p_2\, w_2^{12}, \\ f_3 &= p_3\, u_{33} + \delta\lambda\, \Lambda_3 + p_1^2\, w_3^{11} + p_2^2\, w_3^{22} + 2p_1 p_2\, w_3^{12}, \end{aligned} \tag{4.6}$$

where $\mathbf{w}^{ij} = (1/2)\partial^2 \mathbf{f}/\partial p_i \partial p_j|_{\mathbf{k}_0,\lambda_0}$ $(i,j=1,2)$. We can make a similar transformation on $\mathbf{f}$ such that the f_3 direction is $\mathbf{e}_3 = (\mathbf{u}_3 \times \mathbf{w}^{22})/|\mathbf{u}_3 \times \mathbf{w}^{22}|$, so that f_3 becomes independent of p_3 and p_2^2. Then one also has the freedom to rotate the p_1 and p_2 components to make w_3^{12} vanish. After these two transformations, f_3 only depends on p_1^2 and $\delta\lambda$, and one expects solutions at $p_1 = \pm\sqrt{-\delta\lambda\Lambda_3/w_3^{11}}$. Plugging this into the expressions for f_1 and f_2 in Eq. (4.6), one obtains a quadratic equation for p_2 of the form $a\delta\lambda + bp_2^2 + c\sqrt{\delta\lambda}p_2 = 0$, where a, b and c are some constants determined by the components of $\mathbf{u}_3$, $\mathbf{\Lambda}$, and $\mathbf{w}^{ij}$ $(i,j=1,2)$. If there are real solutions for the above equation, then the doubly-quadratic BTPs would split into four Weyl nodes whose trajectories scale as $p_1 \sim \pm\sqrt{\delta\lambda}$ and $p_2 \sim \pm\sqrt{\delta\lambda}$, $p_3 \sim \delta\lambda$. Otherwise, if there is no solution for p_2, a gap would be opened up immediately after the band touching

at $(\mathbf{k}_0, \lambda_0)$, which would represent the rare case of an "insulator-insulator transition" using the language of Ref. [148].

However, we do not expect that the strong $\mathbb{Z}_2$ index would be inverted for such an insulator-insulator transition in the rank-one case. This can be seen as follows. If the BTP does not lie in any of the TR-invariant slices ($k_j = \{0, \pi\}$, $j = 1, 2, 3$), then certainly the 2D $\mathbb{Z}_2$ indices of the TR invariant slices would not change, and it follows that none of the four 3D $\mathbb{Z}_2$ indices would change either. If the BTP happens to reside in one of the TR invariant slices, then since the dispersion in the 2D slice must be quadratic in at least one direction, it should be topologically equivalent to the superposition of an even number of linearly-dispersing Weyl nodes, which is also not expected to flip the 2D $\mathbb{Z}_2$ index, as argued in Ref. [47]. Thus none of the 3D $\mathbb{Z}_2$ indices, including the strong index, would change.

To summarize this section, we find without any lattice-symmetry restriction that a critical WSM phase always exists in the rank-two case. In the rank-one case, an insulator-insulator type transition is allowed in principle, but would not be expected to be accompanied by a change in the strong $\mathbb{Z}_2$ index. Therefore, it is fair to claim that, regardless of special lattice symmetry, there is always a WSM phase connecting $\mathbb{Z}_2$-odd and $\mathbb{Z}_2$-even phases in a noncentrosymmetric insulator.

4.2.2 Discussion of BiTeI

In this section we discuss the TPT in pressured BiTeI, a case in which the TPT is driven in a system with C_{3v} symmetry. Contrary to the conclusions of Ref. [148], here we argue that a critical WSM does exist in the TPT of BiTeI, although the pressure interval over which it occurs may be rather narrow.

In Refs. [149, 148] the authors argued that if there exists a high-symmetry line in the BZ such that the dispersion extremum evolves along the line as a function of the adiabatic parameter (pressure), then one could get an insulator-insulator type transition

without going through a critical WSM. The authors further pointed out that the high-symmetry lines from A to H in the BZ of BiTeI, shown in Fig. 4.1(d), satisfy some necessary conditions for this to occur. Moreover, they showed that the symmetry of BiTeI is such that if one concentrates on the band dispersions along these A-H lines, one finds a pair of extrema (one valence-band maximum and one conduction-band minimum) which migrate along the A-H line as a function of the external parameter (pressure), coincide at a critical value, and then separate again to reopen the gap. They furthermore showed that the dispersions are quadratic in the two orthogonal directions (except exactly at the critical value), raising the possibility that the extrema in question could be minima and maxima in all three k-space directions. This would correspond to the insulator-insulator transition without an intervening WSM phase. However, our analysis in the previous section shows that this cannot occur in the rank-two case, and that the extrema in question actually become saddle points after the band touchings occur along the A-H lines. In this case, as recognized in Ref. [148], a WSM phase does occur. As has been verified in Ref. [148], the Jacobian does remain of rank two on these lines in BiTeI, and we shall show below in Sec. 4.4.2 that an intermediate WSM phase does occur. We also point out that Fig. 2 of Ref. [148] does not demonstrate the absence of the Weyl nodes, since they are expected to lie off the (k_x, k_z) plane on which the dispersion was plotted.

Yang *et al.* [148] gave another argument in favor of the insulator-insulator scenario in BiTeI as follows. They noted that the band touching first takes place on the A-H line, which is invariant under the combination of time-reversal and mirror operations. This imposes some constraints on the form of the effective Hamiltonian around the BTP, and from these the authors concluded that, if Weyl nodes do appear, they should migrate along trajectories of the form $p_1 \sim \pm\delta\lambda^{1/2}$, $p_2 \sim \pm\delta\lambda^{3/2}$ and $p_3 \sim \delta\lambda$. Such a curve in 3D space possesses non-zero torsion, so that the trajectories of the two Weyl nodes emerging from one quadratic BTP could never join again and form a closed

curve. This implies that if the WSM is formed by such an event, then it would remain permanently, contradicting the fact that BiTeI clearly becomes a globally-gapped TI at higher pressures. Based on this reasoning, they concluded that the TPT in BiTeI must be an insulator-insulator transition without an intermediate WSM.

However, this argument neglects the fact that the C_{3v} symmetry means that there are several A-H lines in the BZ of BiTeI, and the gap first closes by the simultaneous appearance of quadratic BTPs at equivalent positions on all of these lines. Even though the two Weyl nodes which emerge from a single quadratic BTP cannot meet each other, as shown from the torsion of their trajectories, the Weyl nodes from different BTPs can interchange partners and eventually annihilate each other in such a way as to form a closed curve in the BZ. This is exactly the mechanism of the topological phase transition in noncentrosymmetric TIs [46, 47]. As will be discussed in Sec. 4.4.2, there are actually six quadratic BTPs in the full BZ that appear simultaneously, according to the crystalline and TR symmetries. These six Dirac nodes split into twelve Weyl nodes, which are eventually gapped out by annihilation after exchanging partners.

In the following section, we will study the TPTs in various inversion asymmetric materials by first-principles calculations. We predict $LaBi_{1-x}Sb_xTe_3$ and $LuBi_{1-x}Sb_xTe_3$ to be WSM candidates within a certain range of impurity composition x. We also revisit the case of BiTeI, and find that a WSM phase emerges when external pressure is applied to BiTeI, but only within a small pressure interval.

4.3 Preliminaries

4.3.1 Lattice structures and basic topological properties

The assumed crystal structures of $LaBiTe_3$ and $LuBiTe_3$ are very similar to Bi_2Te_3, where five atomic monolayers stack in the [111] direction in an *A*-*B*-*C*-*A*-... sequence forming quintuple layers (QLs) as shown in Fig. 4.1(a). The only difference is that one

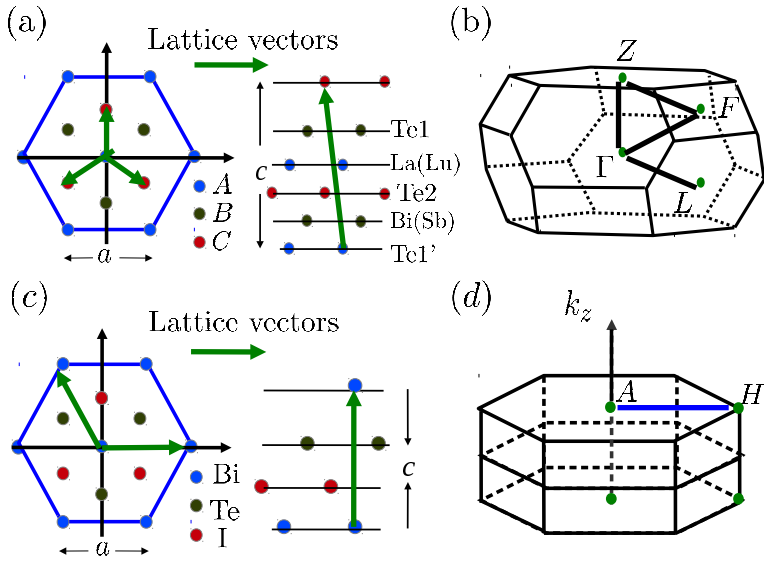

Figure 4.1: (a) The lattice structure of $LaBiTe_3$, $LuBiTe_3$, $LaSbTe_3$, and $LuSbTe_3$. (b) The BZ of La(Lu)Bi(Sb)Te_3. (c) The lattice structure of BiTeI. (d) The BZ of BiTeI.

of the two Bi atoms in the primitive unit cell is replaced by a La or Lu atom, which breaks the inversion symmetry. The lattice structure of $LaSbTe_3$ and $LuSbTe_3$ is the same as for $LaBiTe_3$ and $LuBiTe_3$, except that all the Bi atoms are substituted by Sb. The in-plane hexagonal lattice parameters for $LaBiTe_3$ and $LuBiTe_3$ are $a = 4.39$ Å and 4.18 Å respectively, while the size of a QL along c is 10.07 Å and 10.29 Å respectively. The lattice parameters of $LaSbTe_3$ are slightly different from $LaBiTe_3$, with $a = 4.24$ Å and $c = 10.13$ Å. The lattice parameters for $LuSbTe_3$ have not been reported before, so we use those from $LuBiTe_3$. Among these four hypothetical materials, $LaBiTe_3$ has been previously reported as a candidate for an inversion-asymmetric TI [150]. $LuBiTe_3$ is first reported as a TI candidate in this work; the non-trivial band topology is confirmed by calculating the bulk $\mathbb{Z}_2$ index [130] and checking the existence of topological surface states. On the other hand, $LaSbTe_3$ and $LuSbTe_3$ are trivial insulators.

As shown in Fig. 4.1(c), BiTeI has a hexagonal lattice structure with three atoms in the primitive cell stacked as *A*-*B*-*C*-*A*-... along the z direction. The lattice parameters in-plane and along the hexagonal axis are $a = 4.339$ Å and $c = 6.854$ Å. BiTeI itself is a trivial insulator with a large Rashba spin splitting in the bulk [152], but it can be

driven into a TI state by applying pressure. Previous studies have suggested that the transition to the topological phase is not mediated by a WSM phase [149, 148], but we revisit this issue in Sec. 4.4.2 and come to different conclusions.

4.3.2 First-principles methodology

We carry out the bulk first-principles calculations using the VASP package including SOC [153, 154]. The generalized-gradient approximation is used to treat the exchange-correlation functional [118, 119]. The BZ is sampled on an $8\times8\times8$ Monkhorst-Pack [123] $\mathbf{k}$ mesh and an energy cutoff of 340 eV is used. The output from the first-principles plane-wave calculations are then interfaced to the Wannier90 package [124] to construct realistic tight-binding (TB) models for these materials.[1]

To describe the electronic structure of $LaBi_{1-x}Sb_xTe_3$ and $LuBi_{1-x}Sb_xTe_3$, we adopt the virtual crystal approximation (VCA) in which each Bi or Sb is replaced by a "virtual" atom whose properties are a weighted average of the two constituents. The VCA treatment typically gives a reasonable description for solid-solution systems in which the dopant and host atoms have a similar chemical character. For example, the VCA was shown to work well in describing Sb substitution in Bi_2Se_3, because of the similar orbital character of Sb $5p$ and Bi $6p$, but not for In substitution, where In $5s$ orbitals become involved [42]. The VCA is implemented in the Wannier basis by constructing separate 36-band models for $LaBiTe_3$ ($LuBiTe_3$) and $LaSbTe_3$ ($LuSbTe_3$), including all the valence p orbitals of the cations and anions, as well as the $5d$ and $6s$ orbitals of the rare-earth elements.[2] In the solid solution, the Hamiltonian matrix elements are then taken as a linear interpolation in impurity composition x of the corresponding matrix elements of the parent materials. That is, we take $H_{mn}^{\mathrm{VCA}} = (1-x)H_{mn}^{\mathrm{Bi}} + xH_{mn}^{\mathrm{Sb}}$,

[1] The TB models from Wannier90 are realistic in the sense that the Wannier-interpolated bandstructure can reproduce the first-principles band energies exactly in a specified energy window which, is chosen here to be centered around the Fermi level.

[2] The fully occupied f shell of Lu has very little influence on the electronic structure around the Fermi level.

where H^{Bi}_{mn} and H^{Sb}_{mn} denote the matrix elements of the TB models of $LaBiTe_3$ and $LaSbTe_3$. It worth noting that when generating the WFs for the VCA treatment, the Wannier basis functions have to be chosen as similar as possible before the averaging [42]. We therefore use WFs that are constructed simply by projecting the Bloch states onto the same set of atomic-like trial orbitals without applying a subsequent maximal-localization procedure [52, 53].

Similarly, to study the pressure-induced TPT in BiTeI, we carry out first-principles calculations for the system at the zero-pressure volume, where it is topologically normal, and also at 85.4% of the original volume, a value chosen somewhat arbitrarily to be well inside the TI region [149]. We denote these two states as $\eta = 0$ and $\eta = 1$ respectively. Then from the Wannier representation we again construct a realistic Hamiltonian for each system, denoted as H_0 and H_1 respectively, including all the valence p orbitals of Bi, Te and I. Finally we linearly interpolate these as $H(\eta) = (1 - \eta)H_0 + \eta H_1$, treating η as an adiabatic parameter that tunes the system through the topological phase transition.

Using these Wannierized effective TB models, we can search for BTPs very efficiently over the entire BZ. We first sample the irreducible BZ using a relatively sparse $\mathbf{k}$ mesh, e.g., $20\times20\times20$, and find the point $\mathbf{k}_0$ having the smallest direct band gap on this mesh. A second-round search is conducted by scanning over a denser $\mathbf{k}$ mesh within a sphere centered on $\mathbf{k}_0$. We then repeat the procedure iteratively until convergence is reached. All of the trajectories of Weyl nodes presented in Sec. 4.4 are obtained using this approach.

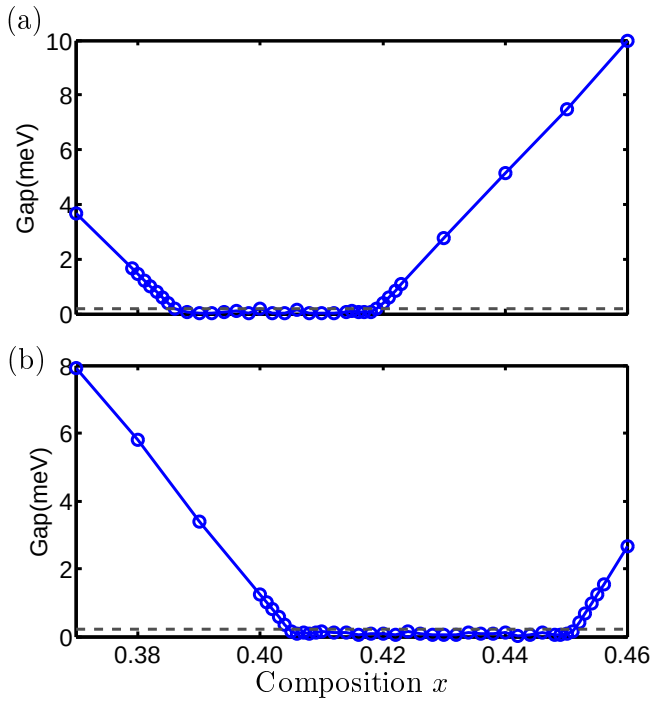

Figure 4.2: Smallest direct band gap in the BZ vs. composition x for (a) $LaBi_{1-x}Sb_xTe_3$ and (b) $LuBi_{1-x}Sb_xTe_3$. Dashed gray line marks our chosen threshold of 0.2 meV to signal a gap closure.

4.4 Results

4.4.1 $LaBi_{1-x}Sb_xTe_3$ and $LuBi_{1-x}Sb_xTe_3$

Band gap and Weyl chirality

For each of these materials we scan over a mesh in composition x, and for each x we construct the Wannierized Hamiltonian for the corresponding solid solution within the VCA. We then use the methods of the previous section to search for the BTPs in the entire irreducible BZ. Plots of the smallest direct band gap in the BZ vs. x are presented in Fig. 4.2. Clearly the gap remains closed over a finite range of x in both cases, from 38.5% to 41.9% for $LaBi_{1-x}Sb_xTe_3$ and 40.5% to 45.1% for $LuBi_{1-x}Sb_xTe_3$. By checking the dispersion around the gap-closure point, we confirm that the system is semimetallic with the Fermi level lying at a set of degenerate Weyl BTPs over this entire range.

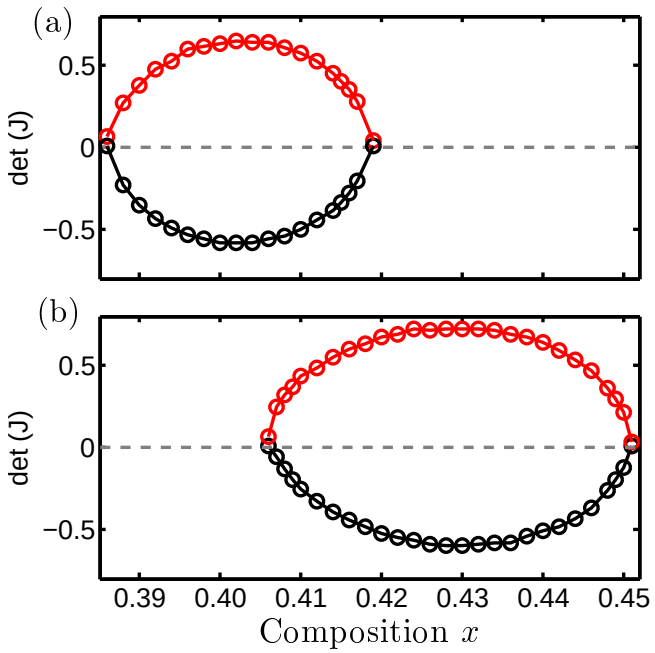

Figure 4.3: Determinant of the Jacobian matrix evaluated at Weyl nodes with positive (red) and negative (black) chirality vs. composition x, for (a) $LaBi_{1-x}Sb_xTe_3$ and (b) $LuBi_{1-x}Sb_xTe_3$.

To illustrate the topological character, we further calculate the chirality of the BTPs, which is given by the determinant of the Jacobian matrix $J_{ij} = \partial f_i/\partial k_j$. Fig. 4.3 shows how det(**J**) varies with x for the BTPs in $LaBi_{1-x}Sb_xTe_3$ and $LuBi_{1-x}Sb_xTe_3$. The red and black open circles mark the values of det(**J**) for the BTPs with positive and negative chirality, which are mapped into each other by mirror operations about the $k_x = 0$ and other equivalent mirror planes. One can see that at the beginning of the band touching, the chirality starts at zero, indicating the creation of a quadratic BTP. As x increases, each quadratic BTP splits into two Weyl nodes with opposite chirality. These then migrate through the BZ and eventually annihilate each other at the point where the chirality returns to zero.

Symmetry considerations

As mentioned earlier, the point group of this class of materials is C_{3v}, which has a 3-fold rotation axis along k_z and three mirror planes that contain the k_z axis and intersect

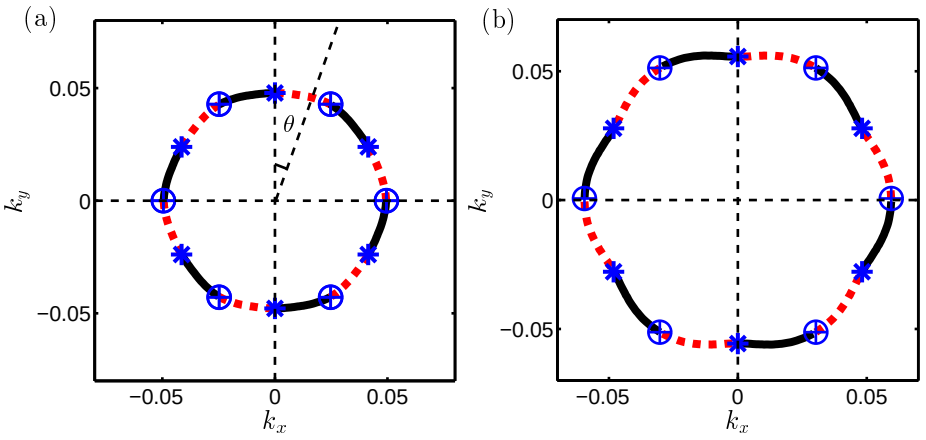

Figure 4.4: Trajectories of Weyl nodes in the (k_x, k_y) plane (in units of Å^{-1}). Dashed red lines indicated Weyl nodes of positive chirality; solid black lines are negative. The '*' and '$\oplus$' denote respectively the points of creation or annihilation of Weyl nodes. (a) For $LaBi_{1-x}Sb_xTe_3$. (b) For $LuBi_{1-x}Sb_xTe_3$.

the $k_z = 0$ plane on the lines $k_x = 0$ and $k_y = \pm k_x/\sqrt{3}$. We define an azimuthal angle θ that measures the rotation of (k_x, k_y) from the $+k_y$ axis in the clockwise direction as shown in Fig. 4.4(a). As a result of the three-fold rotational symmetry, if a Weyl node with positive chirality appears at some θ in the region $0 \leq \theta \leq \pi/3$ and at some k_z in the upper half BZ, then there must be another two nodes with the same chirality and the same k_z located at $\theta + 2\pi/3$ and $\theta - 2\pi/3$. Taking into account the mirror symmetry, these must have negative-chirality partners at the same k_z but at $-\theta$, $-\theta + 2\pi/3$ and $-\theta - 2\pi/3$. Finally, because of TR symmetry, each Weyl node at $\mathbf{k}$ is always accompanied by another at $-\mathbf{k}$ with the same chirality, giving six more Weyl nodes in the lower half BZ. We thus generically expect a total of twelve Weyl nodes in the entire BZ for compositions x in the region of the WSM phase.

Weyl trajectories

Figure 4.4 shows the trajectories of the Weyl nodes in $LaBi_{1-x}Sb_xTe_3$ and $LuBi_{1-x}Sb_xTe_3$ projected onto the (k_x, k_y) plane as x passes through the critical region. The red dashed line represents the trajectory of Weyl nodes with positive chirality, while the solid black

one denotes those with negative chirality, and the "*" and "$\oplus$" denote the creation and annihilation points of the Weyl nodes respectively. As x increases, six quadratic BTPs are simultaneously created in the mirror planes; this occurs at $x_{c1} = 38.5\%$ for $LaBi_{1-x}Sb_xTe_3$ and 40.5% for $LuBi_{1-x}Sb_xTe_3$. Each quadratic BTPs then splits into two Weyl nodes of opposite chirality, and these twelve nodes migrate along the solid black and dashed red lines shown in the figure. Eventually, after exchanging partners, the Weyl nodes meet and annihilate each other in another set of high-symmetry planes ($k_y = 0$ and other equivalent planes), at $x_{c2} = 41.9\%$ for $LaBi_{1-x}Sb_xTe_3$ and 45.1% for $LuBi_{1-x}Sb_xTe_3$.

Figure 4.5(a)-(b) shows the trajectory of the Weyl nodes in the k_z direction. At $x = x_{c1}$, six quadratic BTPs are created, three in the top half-BZ and three in the bottom half-BZ, but all of them fairly close to the BZ boundary plane at $k_z = \pm\pi/c$. As x increases, the six BTPs split to form twelve Weyl nodes, and these begin to move toward the above-mentioned BZ boundary plane. Finally, after interchanging partners, Weyl nodes of opposite chirality annihilate in pairs at x_{c2} on the BZ boundary plane at $k_z = \pm\pi/c$. For $x > x_{c2}$ a global gap opens up and the system is again an insulator but with an inverted $\mathbb{Z}_2$ index.

The locus of Weyl points can be regarded as forming a loop in the 4D space of (k_x, k_y, k_z, x), and just as this loop can be projected onto k_z as in Figs. 4.5(a-b), it can also be projected onto the direction of impurity composition x as shown in Figs. 4.5(c-d). Again, it is clear that the Weyl nodes are created at x_{c1} in the mirror planes and annihilated at x_{c2} at $\theta = \pm\pi/6$. These plots may also be helpful in seeing how the high six-fold symmetry contributes to the narrowness of the WSM region. If the symmetry of the system were lower, the period of oscillation in θ in Figs. 4.5(c-d) would be longer, which would allow the Weyl nodes to oscillate farther in the x direction, giving a wider window of concentration for the WSM phase. In contrast, a fictitious system with an N-fold rotational symmetry would force the width of the WSM region to vanish as

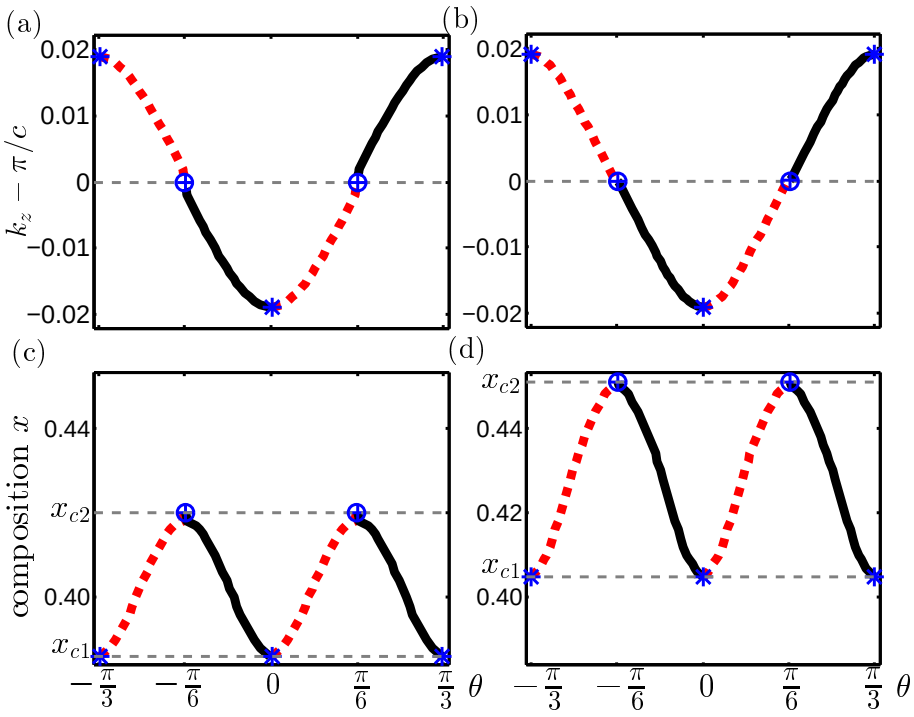

Figure 4.5: (a-b): Trajectories of Weyl nodes in the k_z direction (in units of Å^{-1}) for (a) $LaBi_{1-x}Sb_xTe_3$ and (b) $LuBi_{1-x}Sb_xTe_3$. Dashed red (solid black) lines refer to the Weyl nodes with positive (negative) chirality. θ is the azimuthal angle in the (k_x, k_y) plane, as indicated in Fig. 4.4(a). The "*" and "$\oplus$" denote the creation and annihilation point of the Weyl nodes respectively. (c-d): Trajectories of Weyl nodes in the direction of impurity composition x for (c) $LaBi_{1-x}Sb_xTe_3$ and (d) $LuBi_{1-x}Sb_xTe_3$.

$N \to \infty$. Here we have $N = 6$, which is evidently large enough to limit the WSM phase to a rather small interval in x.

Surface Fermi arcs

One of the most characteristic features of WSMs is the existence of Fermi arcs in the surface bandstructure. Here we calculate the surface states using the surface Green's-function technique [7], which is implemented in the context of the VCA effective Hamiltonian in the Wannier basis. The surface BZ is sampled by a 64×64 **k** mesh, and the surface spectral functions calculated on this mesh are then linearly interpolated to fit a 128×128 **k** mesh. Fig. 4.6 shows the normalized surface spectral functions averaged around the Fermi level for $LaBi_{1-x}Sb_xTe_3$ at $x = 0.405$ and for $LuBi_{1-x}Sb_xTe_3$ at

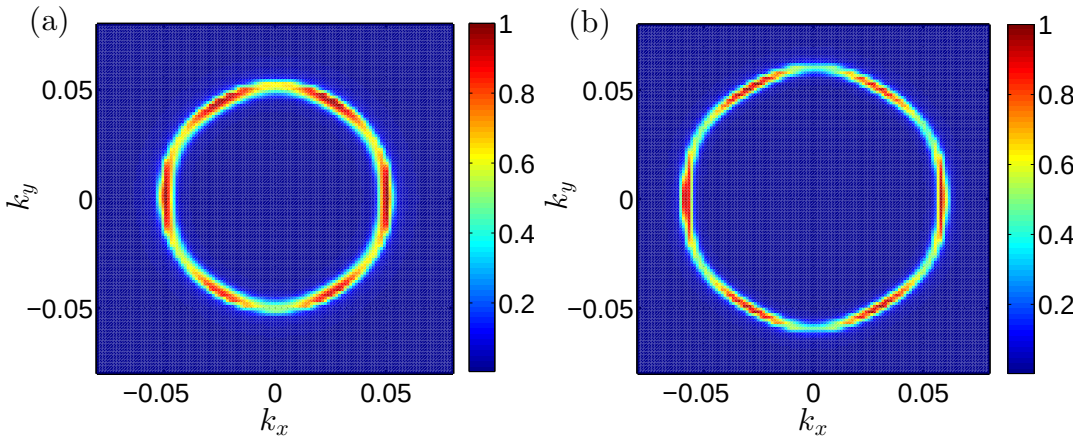

Figure 4.6: Surface spectral function averaged around the Fermi level (k_x and k_y in units of Å^{-1}) for (a) $LaBi_{1-x}Sb_xTe_3$ at $x=0.405$, (b) $LuBi_{1-x}Sb_xTe_3$ at $x=0.43$.

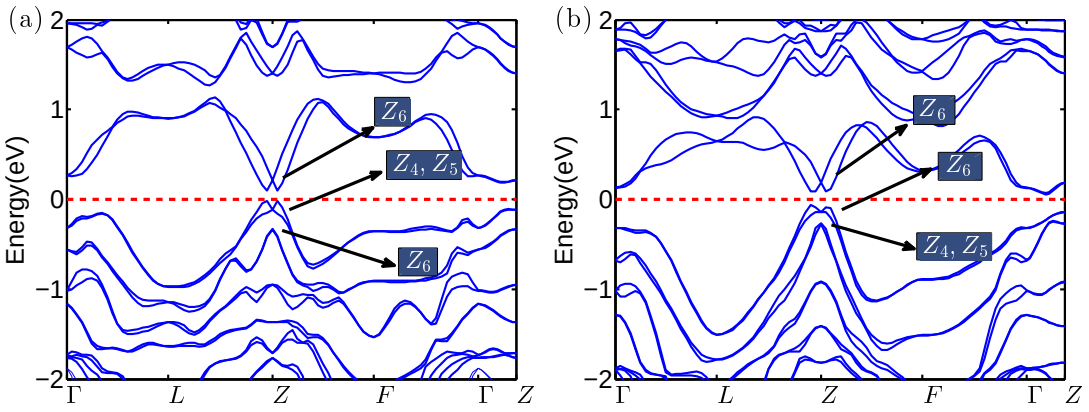

Figure 4.7: Bulk bandstructures of (a) $LaBiTe_3$ and (b) $LuBiTe_3$.

$x=0.43$. The averaging is done over an energy window of ± 4.5 meV around the Fermi energy, which is determined by the position of the bulk Weyl nodes. Six Fermi arcs connecting the projected Weyl nodes of opposite chirality are visible, confirming the existence of the WSM phase in these two solid-solution systems. Note that because of the small projected bulk gap on the loops where the Fermi arcs reside, some non-negligible spectral weight is visible even outside the Fermi arcs in Fig. 4.6, coming from the artificial smearing of the Green's functions.

Simplified six-band model

In order to capture the essential physics in these materials, we construct a six-band TB model to describe the interesting critical behavior. From the bandstructures plots presented in Fig. 4.7, it is clear that the band inversion occurs around the Z point of $LaBiTe_3$ and $LuBiTe_3$, so we focus our attention on the six states at Z closest to the Fermi level. A symmetry analysis shows that these six states belong to two copies of the two-dimensional Z_6 irreducible representation (irrep) of the C_{3v} group at Z, plus a Kramers pair of one-dimensional complex-conjugate Z_4 and Z_5 irreps corresponding to linear combinations of $j_z = \pm 3/2$ orbitals.

We thus build our six-band TB model out of basis states having the symmetry of $|p_z, \uparrow\rangle$ and $|p_z, \downarrow\rangle$ on the Te atoms at the top and bottom of the quintuple layer, and $|p_x + ip_y, \uparrow\rangle$ and $|p_x - ip_y, \downarrow\rangle$ combinations located on the central Te atoms. A schematic illustration of the six-band model is shown in Fig. 4.8, where the top, bottom and central Te atoms are denoted by Te1, Te1′ and Te2 respectively. First of all, six inter-layer spin-independent hopping terms are included in the model. As shown in Fig. 4.8, we consider the first-neighbor hopping between the central and top (bottom) Te atoms t_1 (t_2), the inter-QL (intra-QL) hopping between the top and bottom Te atoms t_3 (t_4), and some further-neighbor hoppings t_u and t_v that are crucial in obtaining a nontrivial $\mathbb{Z}_2$ index. Second, to capture the Rashba spin-splitting in the first-principles bandstructure, in-plane Rashba-like spin-dependent hoppings within the top and bottom Te monolayers are included and are denoted by λ_1 and λ_2 respectively. For completeness, the inter-layer first-neighbor Te1-Te2 (λ_3) and Te1′-Te2 (λ_4) Rashba-like hopping terms are also included. Lastly, to reproduce the first-principles bandstructure better, we also introduce first-neighbor spin-independent hopping terms within the Te1, Te2 and Te1′ monolayers, denoted by v_1, v_2 and v_3 respectively. The onsite energies are also different and are labeled by E_1 for Te1, E_2 for Te2, and E_3 for Te1′. As our model is only intended to be semiquantitative, we use the same model parameters to describe both

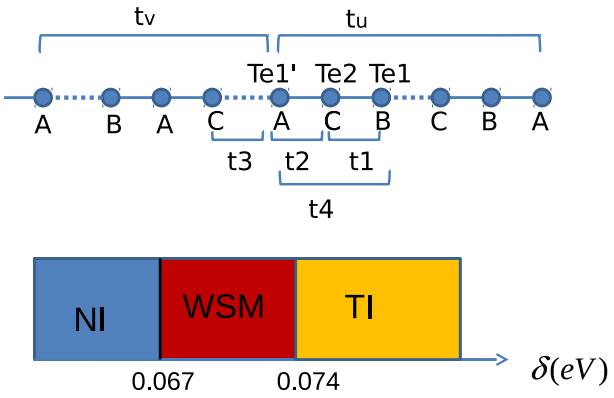

Figure 4.8: Top: Schematic diagram of the inter-layer spin-independent hopping terms in the six-band model. Orbitals on sites Te1, Te2, and Te1′ make up a quintuple layer; A, B and C label in-plane hexagonal positions. Bottom: Phase diagram for the topological behavior of the six-band model.

$LaBiTe_3$ and $LuBiTe_3$.

We take all of the parameters in the model to depend a scaling parameter δ that drives the TPT. When δ is zero, the system is a trivial insulator; as δ increases, the system becomes a topological insulator by going through a critical WSM. The dependence of the parameters on δ defines a path in parameter space. It is important to note that the width of the critical WSM region can be highly sensitive to this path, with an improper choice sometimes leading to an extremely narrow WSM phase. Our choice is specified in Table 4.1.

Table 4.1: Parameters of the six-band model (in eV).

t_1	$0.2-\delta/4$	λ_3	$0.15-\delta/2$	v_3	0
t_2	$0.15-\delta/4$	λ_4	$0.12-\delta/2$	E_1	$0.1+\delta-6v_1$
t_3	δ	t_u	$0.12+\delta/2$	E_2	$-6v_2$
t_4	$0.1-\delta/4$	t_v	$0.06-\delta/2$	E_3	$-0.1-\delta$
λ_1	$0.24-\delta/2$	v_1	0.05		
λ_2	$0.2-\delta/2$	v_2	0.1		

Following the path we have chosen, a WSM phase is obtained for $0.067\,\text{eV} < \delta < 0.074\,\text{eV}$. As shown in Fig. 4.9(b), the smallest direct band gap in the BZ vanishes when $0.067\,\text{eV} < \delta < 0.074\,\text{eV}$, indicating the existence of BTPs in BZ. If one further checks the position of the BTPs, one finds that when $\delta \approx 0.067\,\text{eV}$, six quadratic BTPs are

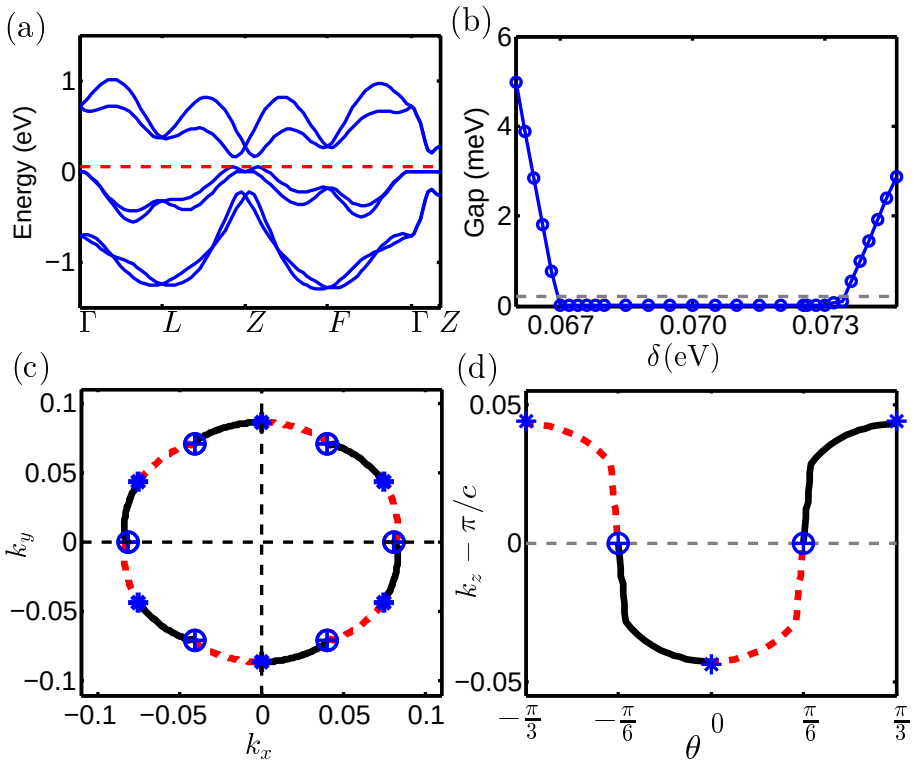

Figure 4.9: (a) Bulk bandstructure of the six-band model at $\delta = 0.09\,\text{eV}$. (b) Smallest direct band gap in the BZ vs. δ. (c) Trajectory of Weyl nodes projected onto the (k_x, k_y) plane. Dashed red (solid black) line refers to the Weyl node with positive (negative) chirality. The "*" and "$\oplus$" denote the creation and annihilation point of the Weyl nodes respectively. θ is the azimuthal angle in the (k_x, k_y) plane. (d) Trajectory of Weyl nodes along k_z. Units of k_x, k_y and k_z are Å^{-1}.

created in the mirror planes, which then split into twelve Weyl nodes and propagate in the BZ following the solid black and dashed red lines in Fig. 4.9(c) and (d). These Weyl nodes eventually annihilate with each other at $\delta \approx 0.074\,\text{eV}$ after exchanging partners, which qualitatively reproduces the phase-transition behavior of the VCA effective Hamiltonians very well. When $\delta > 0.074\,\text{eV}$, the system becomes a strong TI. The bulk bandstructure at $\delta = 0.09\,\text{eV}$ in the TI phase is shown in Fig. 4.9(a), which very well captures the low-energy dispersions around Z that were found in the first-principles calculations.

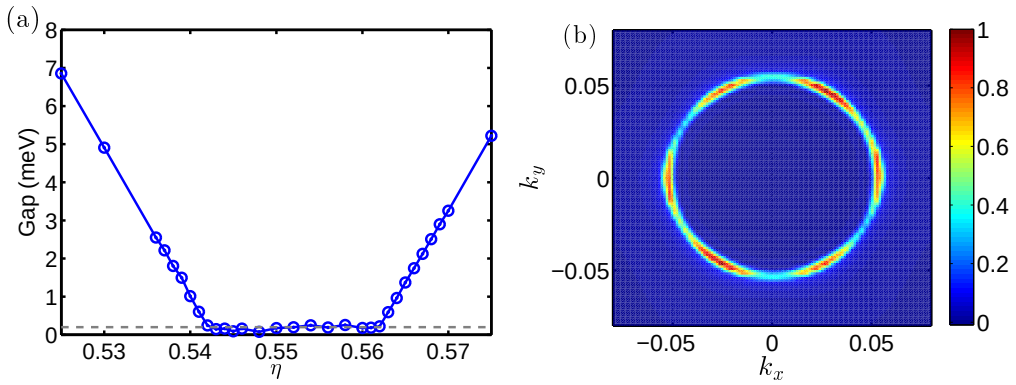

Figure 4.10: (a) Smallest direct band gap in the BZ of BiTeI vs. the pressure-scaling variable η. (b) Surface spectral function of BiTeI in the WSM phase (at 55% of the full pressure).

Discussion

To conclude this section, we would like to comment that the width of the WSM phase depends on two ingredients. On one hand, as discussed above, it depends on the symmetry of the system; other things being equal, the WSM interval tends to be wider in systems with lower symmetry. On the other hand, even for fixed symmetry, it also depends on the the detailed choice of path connecting the topological and trivial phases. Choosing a different path may broaden or reduce the WSM region. For example, if one artificially changes the strength of the atomic SOC strength in $LaBiTe_3$ and $LuBiTe_3$ in the Wannierized TB models, and scales the variation of the actual atomic SOC by a single scaling parameter λ, then we find that the WSM region only shows up for λ in the range of 76.8-77.3%, which is significantly narrower than for the VCA case. However, if an average SOC is applied to the entire system, such that the SOC strength on Te is artificially high and that on Bi is artificially low, we find that a much wider WSM region results. Thus, it may potentially be possible to engineer the width of a critical WSM phase if one can modify the transformation path, as by epitaxial strain, pressure, or additional chemical substitution.

4.4.2 BiTeI: revisited

In order to justify the discussion in Sec. 4.2.2, we revisit the TPT in BiTeI driven by pressure. In our calculations, the pressure is applied by compressing the volume of the primitive cell. The fully compressed volume V is taken to be 85.4% of the original volume V_0, such that the former is well inside the topological region [149], and both the lattice vectors and atomic positions are relaxed at the compressed volume. As discussed in Sec. 4.3.2, we searched for BTPs over the entire irreducible BZ for a transitional Hamiltonian scaled as $H(\eta) = (1-\eta)H_0 + \eta H_1$ for $0 \leq \eta \leq 1$, where H_0 and H_1 represent the Hamiltonians of the uncompressed and fully compressed BiTeI, with even and odd $\mathbb{Z}_2$ indices respectively. As shown in Fig. 4.10(a), as the pressure is increased from 0% to 100% (alternatively, as V is deceased from 100% to 85.4% of V_0), a semimetallic phase emerges for η in the range of about 54-56%.

The point group of BiTeI is the same as for $LaBiTe_3$ and $LuBiTe_3$, namely C_{3v}. Therefore, as explained in Sec. 4.4.1, one would expect the emergence of twelve Weyl nodes in the entire BZ during the phase-transition process. The trajectories of the Weyl nodes are plotted in Fig. 4.11(a-b). When $\eta \approx 54\%$, six quadratic BTPs are first created at the BZ boundary $k_z = \pi/c$ in the $k_y = 0$ and other equivalent high-symmetry planes. These BTPs then split into twelve Weyl nodes which propagate along the directions indicated by solid black (antimonopoles) and dashed red (monopoles) lines. They annihilate each other in the three mirror planes after exchanging partners. Note that in this case the system goes from a normal to topological insulator as η increases, which is the reverse of the $LaBi_{1-x}Sb_xTe_3$ and $LuBi_{1-x}Sb_xTe_3$ cases.

The results shown in Fig. 4.11 support our conclusions in Sec. 4.2.2. In particular, even though the torsion argument implies that the trajectories of the two Weyl nodes which split off from a given quadratic BTP would never meet each other, a closed curve is still formed in the 3D BZ of BiTeI through the interchange of partners among the Weyl nodes.

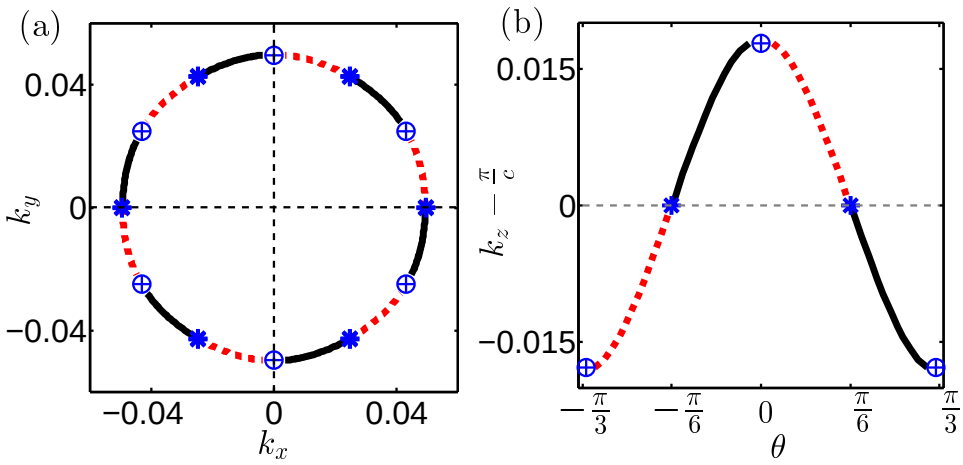

Figure 4.11: (a) Trajectories of Weyl nodes in the (k_x, k_y) plane (in units of Å^{-1}). Dashed red (solid black) lines indicate the trajectories of Weyl nodes with positive (negative) chirality. The "*" and "⊕" denote the creation and annihilation point of the Weyl nodes respectively. (b) Trajectory of Weyl nodes in the k_z direction (units of Å^{-1}).

Fig. 4.10(b) shows the surface spectral function of BiTeI averaged around the Fermi level for $\eta = 0.55$, in the WSM phase. It is clear that there are six Fermi arcs extending between the six pairs of projected Weyl nodes, which is again the hallmark of a WSM phase.

We therefore conclude that a WSM phase does exist in the TPT of BiTeI, but it occurs only within a narrow pressure range. If η is changed by 2.5%, the volume is only changed by 0.39%, which might be difficult to measure experimentally. Again, the narrowness of the WSM interval can be attributed in part to the high symmetry of the system. However, as emphasized in the previous section, the width of the critical WSM is also sensitive to the choice of path in parameter space. The critical WSM could get broadened by choosing a different path, as for example by applying uniaxial pressure. We leave this for a future study.

4.5 Summary

In this chapter, we have investigated the nature of the TPT in a noncentrosymmetric TI in the most general case. We find that an intermediate WSM phase is always present, regardless of other lattice symmetries, as long as inversion symmetry is absent. We discussed separately the cases in which the Jacobian matrix is rank-one or rank-two when the gap first closes. In the rank-two case, each quadratic BTP would always split into a pair of Weyl nodes, which annihilate each other after exchanging partners. If the rank of the Jacobian is one, then the doubly-quadratic BTP in this case would either split into four Weyl nodes, or else immediately be gapped out again, corresponding to an "insulator-insulator transition." However, in the latter case, the bulk $\mathbb{Z}_2$ indices are not expected to change. Therefore, we conclude that $\mathbb{Z}_2$-even and $\mathbb{Z}_2$-odd phases of a noncentrosymmetric insulator must always be separated by a region of WSM phase, even if other symmetries are present.

To illustrate our conclusions, we have carried out calculations on specific noncentrosymmetric insulators. For $LaBi_{1-x}Sb_xTe_3$ and $LuBi_{1-x}Sb_xTe_3$ we have used Wannierized VCA Hamiltonians to find a WSM phase in the region $x \approx 38.5\% - 41.9\%$ and $x \approx 40.5\% - 45.1\%$ respectively. A six-band TB model was also constructed to describe the topological and critical behavior in these materials. We found that the width of the critical WSM phase can be highly sensitive to the choice of path in the parameter space, suggesting that there is flexibility to engineer the WSM phase.

We have also revisited the TPT of BiTeI as a function of pressure, where previous work suggested the absence of a WSM phase [148]. Using a carefully constructed algorithm to search for the minimum gap in the full three-dimensional BZ, we found that a WSM phase is indeed present over a narrow interval of pressure, although this range may be so narrow as to make its experimental observation difficult.

In summary, we have clarified the theory of a general $\mathbb{Z}_2$-even to $\mathbb{Z}_2$-odd topological

phase transition in a three-dimensional time-reversal-invariant insulator with broken inversion symmetry, and demonstrated that an intermediate WSM phase must always be present. We have also detailed the behavior of $LaBi_{1-x}Sb_xTe_3$ and $LuBi_{1-x}Sb_xTe_3$ as promising candidates for WSMs of this kind. While we have not considered disorder or interactions explicitly, we expect our conclusions to survive at least for weak disorder or interactions. Our work is a step forward in the general understanding of topological phase transitions, and may provide useful guidelines for the experimental realization of new classes of Weyl semimentals.

Chapter 5

Spin-orbit spillage as a measure of band inversions in insulators

From the previous chapters, we learned that a crucial concept in topological materials is the so-called "band inversion". Nontrivial band topology is usually associated with an inverted order of bands somewhere in BZ. For example, as discussed above, in Bi_2Se_3, the band inversion driven by spin-orbit coupling takes place at Γ. In this chapter, we deal with such band inversions driven by spin-orbit coupling. We propose a straightforward and effective approach for quantifying the band inversion induced by spin-orbit coupling in band insulators. In this approach we define a quantity as a function of wavevector in the Brillouin zone measuring the mismatch, or "spillage", between the occupied states of a system with and without SOC. Plots of the spillage throughout the BZ provide a ready indication of the number and location of band inversions driven by SOC. To illustrate the method, we apply this approach to the two-band Dirac model, the 2D Kane-Mele model, a 2D Bi bilayer with applied Zeeman field, and to first-principles calculations of some 3D materials including both trivial and $\mathbb{Z}_2$ topological insulators. We argue that the distribution of spillage in the BZ is closely related to the topological indices in these systems. Our approach provides a fresh perspective for understanding topological character in band theory, and should be helpful in searching for new materials with non-trivial band topology.

5.1 Introduction

Spin-orbit coupling (SOC) is a relativistic effect originating from the interaction between the spin and orbital motions of electrons. It has played a key role in various aspects of condensed-matter physics, including the electronic structure of solids and the transport properties in mesoscopic systems [155, 156]. It has been known since the 1950s that SOC can induce anisotropic spin splitting in some III-V semiconductors with the zinc-blende structure, known as the Dresselhaus splitting [155]. In 2D and quasi-2D systems, the SOC resulting from the electric field perpendicular to the 2D plane gives rise to a Rashba splitting linear in k with interesting "helical" spin textures [155, 156]. The SOC is also crucial in determining the transport behavior of low-dimensional electronic systems. One famous example is the weak antilocalization in spin-orbit-coupled 2D electronic systems, where the backscattering amplitudes interfere destructively due to a geometric Berry phase [11] associated with the intrinsic SOC, leading to a suppressed resistivity when an external magnetic field is absent [157]. SOC is also responsible for spin precession in 1D and quasi-1D systems [156], the spin Hall effect in paramagnetic metals [158], and numerous other effects.

The SOC has received renewed attention recently because of its central role in the physics of topological insulators (TIs) and related topological states. Typically, the transition from a topologically trivial to a non-trivial phase is accomplished by a SOC-driven inversion of states of different symmetry at the conduction-band minimum (CBM) and valence-band maximum (VBM). For example, such a SOC-driven topological band inversion between Γ_6-derived (s-like) and Γ_8-derived (p-like) states at the zone center is responsible for the quantum spin Hall (QSH) state observed in HgTe/CdTe quantum wells [28, 89]. Similarly, the Kane-Mele model of 2D graphene-like systems [27, 26] enters the QSH state when two band inversions occur at the K and K$'$ points as the SOC strength is increased at a constant staggered potential. In 3D band insulators with time-reversal (TR) symmetry, a SOC-induced band inversion can transform the

system from a trivial insulator into a strong TI displaying an odd number of gapless Dirac cones in the surface states, as occurs for Bi_2Se_3 and Bi_2Te_3 [33, 34, 111, 36].

In the case of a 3D strong TI with inversion symmetry such as Bi_2Se_3, the strong $\mathbb{Z}_2$ index can be uniquely determined by the parities of the occupied bands at the TR-invariant momenta (TRIM) in the Brillouin zone (BZ) [90]. If the highest occupied states and lowest unoccupied states at one of the TRIM possess opposite parities without SOC, and they are inverted by turning on SOC, then the system transforms from a normal to a topological insulator. For example, in Bi_2Se_3, two pairs of Kramers-degenerate occupied states at the BZ center (Γ) are inverted by SOC, resulting in the nontrivial $\mathbb{Z}_2$ index. For TIs without inversion symmetry, the band inversion may happen at arbitrary points in the BZ, instead of at the TRIM. We can identify such band inversion points as the points where a band touching occurs between valence and conduction bands as the SOC is adiabatically turned on; TR symmetry implies that an inversion at $\mathbf{k}_0$ will always be accompanied by one at $-\mathbf{k}_0$. Even in the absence of inversion symmetry, therefore, a band inversion driven by SOC is typically a hallmark of the non-trivial topology in TIs with TR symmetry.

The SOC also plays a crucial role in giving rise to the Chern insulator (CI) state, also known as the quantum anomalous Hall state, which can occur in 2D insulators lacking time-reversal symmetry. The possibility of a CI state was first introduced by Haldane [24], who constructed an explicit model that demonstrates the effect. Although the Haldane model is a model of spinless Fermions on a honeycomb lattice, its key feature is the presence of complex second-neighbor hoppings, which can be regarded as arising from intrinsic atomic SOC through a second-order perturbation process in a more realistic system of spinor electrons [159]. An example is a Bi bilayer with an applied Zeeman field, as will be discussed below.

The concept of topological band inversion has been much discussed in the topological-insulator literature, but in the absence of symmetry it may be difficult to recognize when

a band inversion has actually occurred. The usual approach is to look at the symmetry or orbital character at a high-symmetry point in the BZ where a band inversion is suspected, but this only works if sufficient symmetry is present. Some authors have tried to deduce the presence of band-inversion behavior by studying other properties of the system, such as by looking at the qualitative shape of the bands near the symmetry point [160], or even more indirectly, by studying the variation of the band-energy differences with strain in the absence of SOC [161]. However, the reliability of such methods is questionable, as they do not give a direct and quantitative evaluation of the SOC-induced band inversion.

In this chapter, we propose that the calculation of spin-orbit spillage, which measures the degree of mismatch between the occupied band projection operators with and without SOC, provides a simple and effective measure of SOC-driven band inversion in insulators. We demonstrate that the mapping of this spin-orbit spillage in k-space easily allows a direct identification of any region in the BZ where band inversion has occurred, and that the maximum spillage is a useful indicator of topological character. We illustrate the method in the context of both tight-binding models and realistic first-principles calculations.

This chapter is organized as follows. In Sec. II the formal definition of SOC-induced spillage is introduced, and the correspondence between topological indices and spillage is also discussed. In Sec. III the formalism is applied to various systems, including the two-band Dirac model, 2D Kane-Mele model, a Bi bilayer with tunable SOC and exchange field, and realistic materials including Bi_2Se_3, In_2Se_3, and Sb_2Se_3. In Sec. IV we make a brief summary.

5.2 Formalism

5.2.1 Definitions

Mathematically, the mismatch between two projection operators P and $\widetilde{P}$, both of rank N, can be represented by a quantity

$$\gamma = N - \mathrm{Tr}[P\widetilde{P}] = \mathrm{Tr}[P\widetilde{Q}] = \mathrm{Tr}[Q\widetilde{P}] \tag{5.1}$$

where $Q = 1 - P$ and $\widetilde{Q} = 1 - \widetilde{P}$ denote the complementary projections. This measure of mismatch is often referred to as "spillage" since it measures the weight of states that spill from P into $\widetilde{Q}$, or equivalently, from $\widetilde{P}$ into Q. Clearly the spillage vanishes if $P = \widetilde{P}$ at one extreme, and rises to N at the other extreme if there is no overlap at all between the subspaces associated with P and $\widetilde{P}$. Thus, the spillage provides a measure of the degree of mismatch between the two subspaces.

Here we apply this concept to the band projection operators

$$P(\mathbf{k}) = \sum_{n=1}^{n_{\mathrm{occ}}} |\psi_{n\mathbf{k}}\rangle\langle\psi_{n\mathbf{k}}| \tag{5.2}$$

associated with a given wavevector $\mathbf{k}$ in the BZ of a crystalline insulator with $N = n_{\mathrm{occ}}$ occupied bands. We assume an effective single-particle Hamiltonian such as that appearing in density-functional theory (DFT) [3, 4]. Then the SOC-induced spillage $\gamma(\mathbf{k})$ is defined as

$$\gamma(\mathbf{k}) = \mathrm{Tr}[P(\mathbf{k})\widetilde{Q}(\mathbf{k})] \tag{5.3}$$

where P and $\widetilde{P}$ (and their complements) refer to the system with and without SOC respectively. More explicitly,

$$\begin{aligned}\gamma(\mathbf{k}) &= n_{\mathrm{occ}} - \mathrm{Tr}[P(\mathbf{k})\widetilde{P}(\mathbf{k})] \\ &= n_{\mathrm{occ}} - \sum_{m,n=1}^{n_{\mathrm{occ}}} |M_{mn}(\mathbf{k})|^2\end{aligned} \tag{5.4}$$

where

$$M_{mn}(\mathbf{k}) = \langle\psi_{m\mathbf{k}}|\tilde{\psi}_{n\mathbf{k}}\rangle \tag{5.5}$$

is the overlap between occupied Bloch functions with and without SOC at the same wavevector $\mathbf{k}$. Equivalently, this can be written as $M_{mn}(\mathbf{k}) = \langle u_{m\mathbf{k}} | \tilde{u}_{n\mathbf{k}} \rangle$ if one prefers to work in terms of the cell-periodic $|u_{n\mathbf{k}}\rangle$ defined as $u_{n\mathbf{k}}(\mathbf{r}) = e^{-i\mathbf{k}\cdot\mathbf{r}}\psi_{n\mathbf{k}}(\mathbf{r})$.

In the case of realistic DFT calculations in a plane-wave basis, the overlap matrix elements are easily evaluated as

$$M_{mn}(\mathbf{k}) = \sum_{\mathbf{G}} \langle \psi_{m\mathbf{k}} | \mathbf{k}+\mathbf{G} \rangle \langle \mathbf{k}+\mathbf{G} | \tilde{\psi}_{n\mathbf{k}} \rangle \,, \tag{5.6}$$

where $|\mathbf{k}+\mathbf{G}\rangle$ is the plane wave $e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}$ for reciprocal vector $\mathbf{G}$ normalized to the unit cell. The evaluation should also be straightforward in other first-principles basis sets. For simple lattice models the Hamiltonian is typically written in an orthonormal tight-binding basis, so that the wavefunctions are

$$|\psi_{n\mathbf{k}}\rangle = \sum_{j} C_{nj,\mathbf{k}} \, |\chi_{j\mathbf{k}}\rangle \tag{5.7}$$

where $|\chi_{j\mathbf{k}}\rangle$ are the Bloch basis states

$$\chi_{j\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \, \varphi_j(\mathbf{r}-\mathbf{R}) \tag{5.8}$$

and $\varphi_j(\mathbf{r}-\mathbf{R})$ is the j'th tight-binding basis orbital in unit cell $\mathbf{R}$. Then the spillage is trivially computed using

$$M_{mn}(\mathbf{k}) = \sum_{j} C^{*}_{mj,\mathbf{k}} \widetilde{C}_{nj,\mathbf{k}} \,. \tag{5.9}$$

Since the use of Wannier interpolation methods [53, 10, 9] is becoming increasingly frequent, we also comment on this case here. In this approach, the occupied Bloch states are again written as in Eq. (5.7), but this time the Bloch basis states are

$$\chi_{j\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \, w_j(\mathbf{r}-\mathbf{R}) \tag{5.10}$$

where $w_j(\mathbf{r}-\mathbf{R})$ is the j'th Wannier function in unit cell $\mathbf{R}$. Then the spillage is again computed using Eqs. (5.4) and (5.9). This will be accurate as long as the WFs for the systems with and without SOC are chosen to be the same, or as similar as possible. As

we shall see in the following section, the results from the Wannier basis match those of the direct plane-wave calculation very closely for the cases studied here.

In the case of complex unit cells or supercells with many bands near the gap, it may be difficult to identify precisely which bands have been inverted by the SOC. In this case it may be helpful to define a valence-band-resolved spillage as $\gamma_n(\mathbf{k}) = [L(\mathbf{k})L^\dagger(\mathbf{k})]_{nn}$, where $L_{nm}(\mathbf{k}) = \langle \psi_{n\mathbf{k}} | \tilde{\psi}_{m\mathbf{k}} \rangle$ is the overlap matrix between the occupied states without SOC and the unoccupied states with SOC. Then the total spillage is $\gamma(\mathbf{k}) = \sum_{n=1}^{n_{\mathrm{occ}}} \gamma_n(\mathbf{k})$. Similarly, $\bar{\gamma}_m = (L^\dagger L)_{mm}$ provides a conduction-band-resolved spillage. However, it should be noted that γ_n and $\bar{\gamma}_m$ are not gauge-invariant; they will change under a unitary transformation among the occupied or unoccupied states. A natural gauge choice is the one associated with the singular-value decomposition $L = V\Sigma W^\dagger$. Transforming the sets of occupied and unoccupied states according to the unitary matrices V and W respectively, the overlap matrix between the transformed states is just Σ, which is real and diagonal. The columns of V (W) corresponding to the leading eigenvalues indicate which linear combinations of valence (conduction) states contribute the most to the total spillage. We leave the exploration of these refinements for a future study.

5.2.2 Relation to topological character

Here we argue that the presence of non-trivial topological indices will be reflected in certain features of the spillage distribution in the BZ.

We first consider the relatively simple case in which the SOC-driven band inversion is associated with the crossing of highest valence and lowest conduction states belonging to two different irreducible representations (irreps) at a high-symmetry point $\mathbf{k} = \Lambda_0$ in the BZ. Since the states belonging to different irreps have no overlap with each other, the spillage at Λ_0 must be greater than or equal to the irrep dimension. In TR-invariant Bi_2Se_3, for example, the four states around the Fermi level at Γ consist of two Kramers

doublets of opposite parity. In this case the dimension of the irreps is two, so we expect a peak in $\gamma(\mathbf{k})$ centered at Γ whose height is $\gamma_{\max} \geq 2$. As we shall show in Sec. 5.3.4, this is exactly what we find in Bi_2Se_3.

Next, we argue that a correspondence between topological character and spillage should also remain valid for more general cases without special lattice symmetry. Let us first consider the case of CIs (i.e., with broken TR symmetry). We assume the Bloch functions $\psi_{n\mathbf{k}}$ are those of a normal system with Chern number $C = 0$, while $\tilde{\psi}_{n\mathbf{k}}$ are topologically nontrivial with a nonzero Chern number $\widetilde{C}$. We argue that this implies the existence of at least one point in the BZ where the spillage is ≥ 1. If we assume the contrary, i.e., $\gamma(\mathbf{k}) < 1$ everywhere in the BZ, then the determinant of the overlap matrix of Eq. (5.5) between $\psi_{n\mathbf{k}}$ and $\tilde{\psi}_{n\mathbf{k}}$ obeys $\det(M_{\mathbf{k}}) > 0$ everywhere, since a singular M would imply $\gamma \geq 1$. Because the system $|\psi_{n\mathbf{k}}\rangle$ is topologically normal, we know it is possible to choose a smooth and periodic gauge for it, and we assume without loss of generality that this has been done. But if $M_{\mathbf{k}}$ is nowhere singular, the $|\psi_{n\mathbf{k}}\rangle$ can be used as "trial functions" to construct a smooth and periodic gauge for the $|\tilde{\psi}_{n\mathbf{k}}\rangle$, as follows. At each $\mathbf{k}$, carry out a singular value decomposition to express $M = V^\dagger \Sigma W$ (V and W are unitary and Σ is real positive diagonal), and then use the unitary matrix $V^\dagger W$ to transform the original $\tilde{\psi}_{n\mathbf{k}}$ to a new set $\tilde{\psi}'_{n\mathbf{k}}$. Then $M = V^\dagger \Sigma V$, i.e., it is Hermitian and positive definite. Intuitively, this means that a smooth and periodic gauge has been chosen for the states $\tilde{\psi}'_{n\mathbf{k}}$ to make them "maximally aligned" with the states $\psi_{n\mathbf{k}}$. But a smooth and periodic gauge is inconsistent with a nonzero Chern number, completing the proof by contradiction. Thus, if $\gamma < 1$ everywhere in the BZ, then $M_{mn,\mathbf{k}}$ is nonsingular everywhere, and the system $|\tilde{\psi}_{n\mathbf{k}}\rangle$ is normal. Conversely, a topological system must have $\gamma(\mathbf{k}) \geq 1$ somewhere in the BZ, which provides both a signal for the topological phase and an indication of where in the BZ the band inversion has occurred.

For the TR-invariant $\mathbb{Z}_2$ TIs, similar arguments can be put forward that work even

in the absence of inversion symmetry. If the system of $|\psi_{n\mathbf{k}}\rangle$ is in the $\mathbb{Z}_2$-even phase, one can always make a smooth gauge choice over the entire BZ that respects TR symmetry. In the $\mathbb{Z}_2$-odd case, however, such a gauge choice does not exist [56, 57]. Therefore, $\det(M_{\mathbf{k}})$ must vanish somewhere in BZ, or else the smooth gauge could be transferred to the $|\tilde{\psi}_{n\mathbf{k}}\rangle$, resulting in a contradiction. Due to the TR symmetry, $\det(M_{\mathbf{k}}) = \det(M_{-\mathbf{k}})$, so one would generically expect $\gamma(\mathbf{k}) \geq 1$ at two points ($\mathbf{k}_0$ and $-\mathbf{k}_0$) in the BZ. For the case of inversion-symmetric TIs, $\mathbf{k}_0$ and $-\mathbf{k}_0$ merge at one of the TRIM, the two spillages add up, and one expects $\gamma \geq 2$ at one of the TRIM.

In the following section, we numerically test and confirm the above arguments by applying the formalism to systems in different topological phases.

5.3 Applications

5.3.1 Application to two-band Dirac Hamiltonian

As a warm-up exercise, we first apply the spillage formula to a minimal model of a band inversion in 2D (k_x, k_y) space, namely a Dirac model at half filling as described by the Hamiltonian

$$H = m(1-\lambda)\sigma_z + k_x\sigma_x + k_y\sigma_y \tag{5.11}$$

where σ_j are Pauli matrices. Here m is a mass and λ is a control parameter that inverts the bands at $\lambda = 1$. Physically, such a model may describe the low-energy physics in the vicinity of a band touching event associated with the transition from a normal to a quantum anomalous Hall insulator, or at one of the band touching events (at $\mathbf{k}_0$ or $-\mathbf{k}_0$) in the transition to a spin-Hall insulator. The energy spectrum of the above Hamiltonian is $E_{\pm} = \pm\sqrt{m^2(1-\lambda)^2 + k_x^2 + k_y^2}$, where the gap closes at $\lambda = 1$ at Γ ($k_x = k_y = 0$). The spillage is just $\gamma(\lambda, \mathbf{k}) = 1 - |\langle\psi^0_{1\mathbf{k}}|\psi^{\lambda}_{1\mathbf{k}}\rangle|^2$, where $|\psi^0_{1\mathbf{k}}\rangle$ ($|\psi^{\lambda}_{1\mathbf{k}}\rangle$) is the occupied eigenstate at zero (non-zero) λ.

Figure 5.1 shows the spillage *vs.* k_x at $k_y = 0$ as λ is increased from 0.4 to 1.9. When

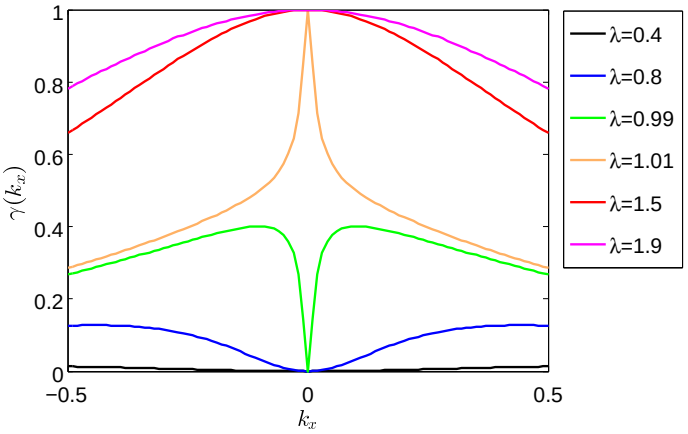

Figure 5.1: The spillage of the half-filled Dirac Hamiltonian as λ increases from 0.4 to 1.9.

$\lambda = 0.4$, the spillage is negligible almost everywhere, and is exactly zero at Γ. On the other hand, when $\lambda = 0.99$, which is very close to the gap closure point, one finds two peaks of spillage emerging on either side of Γ, with a peak value approaching 0.5 as $\lambda \to 1$. As λ passes through the critical point at $\lambda = 1$, one finds that the spillage at Γ jumps from 0 to 1, and then gradually spreads out in BZ as λ is increased further.

This interesting behavior can be interpreted as follows. When $\lambda = 0$, the σ_z term dominates around Γ, so that the pesudospin is mostly along the z direction around Γ. On the other hand, if λ is very close to 1, the σ_x and σ_y terms dominate near (but not exactly at) Γ, forcing the pseudospin direction to point in the (x, y) plane and resulting in a spillage of 1/2. However, the σ_x and σ_y terms vanish at Γ, which means the pseudospin has to point along the $\pm z$ direction. Therefore, when $\lambda < 1$ ($\lambda > 1$), the pseudospin is parallel (anti-parallel) with the pseudospin direction at $\lambda = 0$, such that the spillage jumps from 0 to 1 as λ passes through the critical point.

5.3.2 Application to the Kane-Mele model

The Kane-Mele model is a four-band TB model on a graphene lattice, including nearest-neighbor (NN) spin-independent hoppings and both NN and next-NN spin-dependent

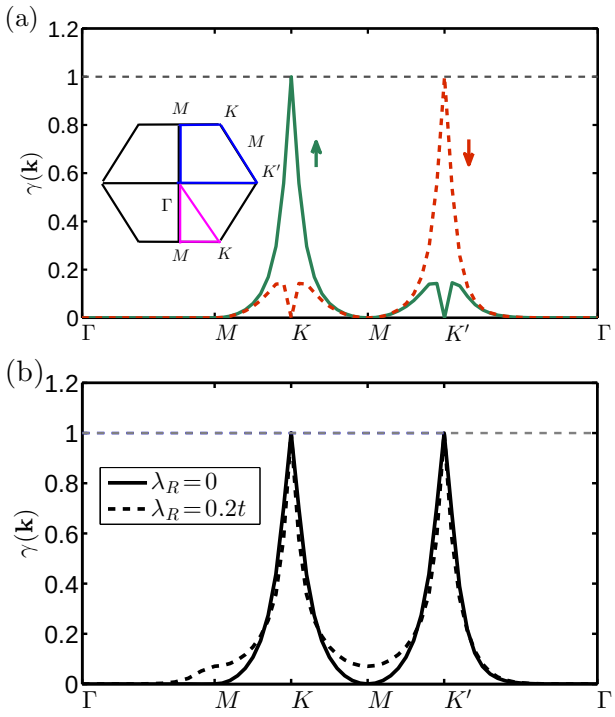

Figure 5.2: Spin-orbit spillage of the Kane-Mele model in the QSH phase, with $t=1$, $\lambda_{\rm so}=0.1t$, and $\epsilon=0.1t$. (a) Spin-resolved spillage without Rashba coupling; solid (green) and dashed (red) lines denote spin-up and spin-down spillage. Inset shows Γ-M-K-M-K′-Γ path used here (blue) and K-Γ-M-K path used in Fig. 5.3 (magenta). (b) Total spillage without (solid line) and with (dashed line) Rashba coupling.

hoppings:

$$H=\sum_{\langle ij\rangle} tc_i^\dagger c_j+\sum_{\langle\langle ij\rangle\rangle} i\lambda_{\rm so}\nu_{ij}c_i^\dagger s_z c_j+\sum_{\langle ij\rangle} i\lambda_{\rm R}c_i^\dagger(\mathbf{s}\times\hat{\mathbf{d}}_{ij})_z c_j+\sum_i \epsilon(-1)^i c_i^\dagger c_i\,. \qquad (5.12)$$

Here spin is implicit, t is the NN spin-independent hopping amplitude, $\lambda_{\rm so}$ is the strength of the next-NN non-spin-flip SOC, $\lambda_{\rm R}$ is the NN Rashba-like SOC amplitude, and ϵ is the magnitude of on-site energy, with signs ±1 for A and B sublattices respectively. Also, $\nu_{ij}=\pm1$ with the sign depending on the chirality of the next-NN bond from site i to j, and $\hat{\mathbf{d}}_{ij}$ is the unit vector pointing from site i to its NN j. In this model, $\lambda_{\rm so}$ competes with $\lambda_{\rm R}$ and ϵ, in the sense that $\lambda_{\rm so}$ tends to drive the system to the QSH phase while $\lambda_{\rm R}$ and ϵ tend to retain the trivial band topology.

For simplicity, we first drop the Rashba coupling, so that spin is a good quantum number. The system is in the QSH phase when $3\sqrt{3}\lambda_{\rm so}>\epsilon$, and in the normal phase

otherwise. Without the Rashba term, the Kane-Mele model can be considered as a superposition of two copies of the Haldane model with opposite Chern numbers [24]. If one calculates the 2D Chern numbers for spin-up and spin-down electrons separately, one would find that the two Chern numbers are ± 1 in the QSH phase. While the Haldane-model system goes from a normal insulator to a CI via a band inversion at either the K or K$'$ point, the Kane-Mele model transitions to the QSH state via simultaneous band inversions at both K and K$'$, but for opposite spins at these two points.

The SOC-induced spillage without the Rashba term is shown in Fig. 5.2(a). In this case the spins act independently, so the spin-up and spin-down spillages $\gamma_\sigma(\mathbf{k}) = n_{\mathrm{occ}}/2 - \sum_{m,n=1}^{n_{\mathrm{occ}}/2} |M_{n\sigma,m\sigma}(\mathbf{k})|^2$ (where $\sigma = \{\uparrow,\downarrow\}$) are shown separately. Clearly the spin-up band inversion at K is responsible for $\gamma_\uparrow = 1$, and conversely at K$'$. The total spillage $\gamma(\mathbf{k}) = \gamma_\uparrow(\mathbf{k}) + \gamma_\downarrow(\mathbf{k})$ is shown by the solid line in Fig. 5.2(b). The symmetry between the behavior at K and K$'$ has been restored by summing over spins. Note that the peak values are $\gamma = 1$ exactly; the fact that they do not exceed one is an artifact of the simplicity of the model. It is also interesting to note that in the absence of time-reversal symmetry, the spin-resolved spillage is closely related to the van Vleck paramagnetism in spin-orbit coupled systems.

When the Rashba coupling is included, as shown by the dashed line in Fig. 5.2(b), spin is no longer a good quantum number, so that a spin decomposition is not well-defined. As expected, adding the Rashba term does not significantly change the results [1]; one still finds that the spillage reaches unity at K and K$'$ as before, providing an indication of the spin-Hall phase.

5.3.3 Application to Chern insulators

We now consider the case of broken TR symmetry, so that the $\mathbb{Z}_2$ index is no longer well-defined, but the possibility of CI phases appears. As discussed in Sec. 5.1, SOC is

[1] The $\mathbf{k}$-dependence of the Rashba coupling terms is such that they vanish at K and K$'$.

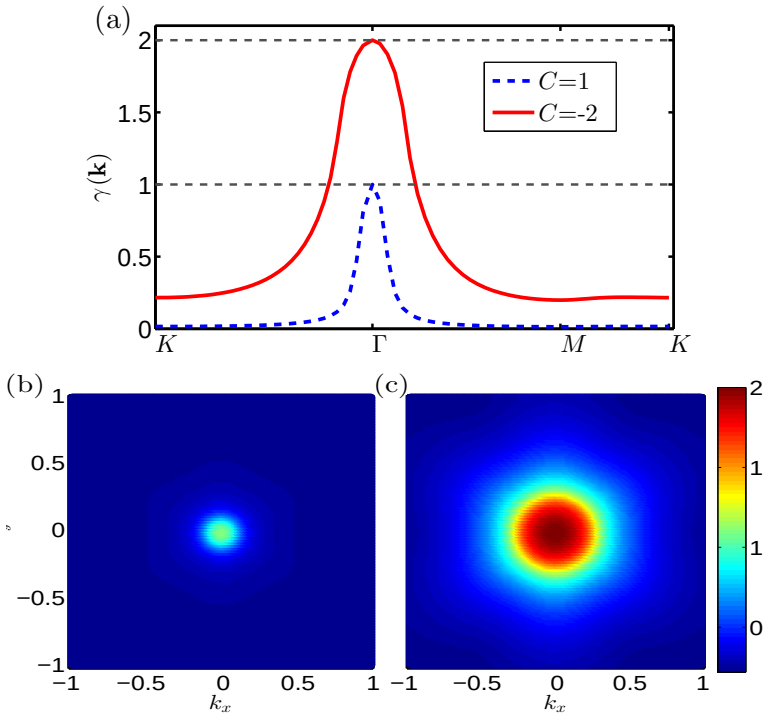

Figure 5.3: (a) Spin-orbit spillage of the Bi bilayer for $C=1$ (dashed blue) and $C=-2$ (solid red) phases, plotted along the K-Γ-M-K path (magenta path in inset of Fig. 5.2(a)). (b) Spillage for $C=1$ phase plotted in the 2D BZ (k_x and k_y in units of Å^{-1}). (c) Same for $C=-2$ phase.

important here as well. Here we study a buckled honeycomb Bi bilayer with a Zeeman field applied normal to the plane, which can be regarded as having been cut from a 3D Bi crystal on a (111) plane. The Bi (111) bilayer has been proposed as a candidate for QSH insulator [91]. If a Zeeman field is further applied, it is possible to obtain CI phases with Chern numbers $C=1$ or $C=-2$ [162, 159]. To describe this system we use a TB model based on Bi $6s$ and $6p$ orbitals, where the first-neighbor ss, sp, $pp\sigma$, and $pp\pi$ hoppings, as well as the second-neighbor $pp\sigma$ hoppings, are included. The hopping parameters are taken from a TB model for 3D bulk Bi [163]. In order to obtain non-zero Chern numbers, an on-site p-shell SOC (λ_{SOC}) and a Zeeman field (H_z) are further applied. It turns out that if H_z is fixed at 0.8 eV, then the phases with $C=-2$ and $+1$ are realized at $\lambda_{\text{SOC}}=2.4$ eV and 0.6 eV respectively. If the SOC is completely turned off, $C=0$.

The spillage for the Bi bilayer is shown along a high-symmetry k-path in Fig. 5.3(a), and as a distribution in the 2D BZ in Figs. 5.3(b-c), for the two parameter sets giving the $C = 1$ and $C = -2$ phases. In both cases the spillage distribution is concentrated at Γ, indicating a band inversion there, although it is much more sharply peaked in the $C = 1$ case. Clearly the spillages provide a signature of the presence of a Chern-insulator phase, including the location of the band inversion and the magnitude (but not the sign) of the Chern number. Here again the peak values of the spillage are exactly equal to the magnitude of the Chern number. For more realistic systems with more bands included, the spillage can be expected to exceed these values slightly, but a clear correlation between the peak values of spillage and the Chern number is still expected.

5.3.4 Application to 3D topological insulators

In this subsection we apply our formalism to realistic first-principles calculations of Bi_2Se_3, In_2Se_3 and Sb_2Se_3. Bi_2Se_3 is a well-known strong TI [36], where the SOC-induced band inversion takes place at Γ. We also consider In_2Se_3 and Sb_2Se_3 in the same crystal structure (known as β phase for In_2Se_3 and not realized experimentally for Sb_2Se_3), which are theoretically predicted (and experimentally confirmed for In_2Se_3) to be trivial insulators [36, 41, 40, 42]. Here it is interesting to note that even though Sb and In have very similar atomic SOC strength, the substitution of In atoms tends to drive Bi_2Se_3into a trivial-insulator phase much faster than does Sb substitution, due to the existence of In $5s$ orbitals [42].

As shown in Fig. 5.4, the considered structure is rhombohedral, with two cations and three Se atoms in the primitive unit cell. The five 2D monolayers are stacked in an A-B-C-A-... sequence along the (111) direction to form quintuple layers (QLs). Experimentally the in-plane hexagonal parameters are $a = 4.138$ and 4.05 Å, and the QL size is $c = 9.547$ and 9.803 Å, for Bi_2Se_3 and In_2Se_3 respectively. In our calculations, we

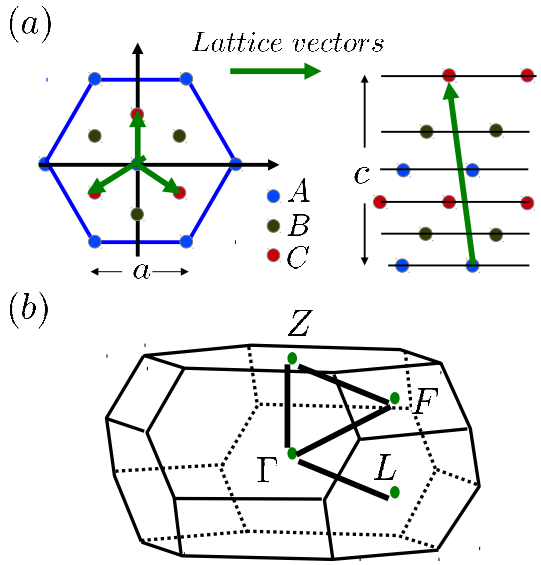

Figure 5.4: (a) Lattice structure of Bi_2Se_3. (b) The BZ of Bi_2Se_3; the spillage and bandstructures shown in Fig. 5.5(a) and Fig. 5.6 are plotted along the black path.

take the experimental lattice parameters of Bi_2Se_3 and In_2Se_3, but relax their internal atomic coordinates. As for Sb_2Se_3, because its rhombohedral structure is not adopted in nature, both the lattice parameters and atomic positions are relaxed. The ground state of rhombohedral Sb_2Se_3 is predicted to be a trivial insulator with $a = 4.11$ Å and $c = 10.43$ Å.

We use the QUANTUM ESPRESSO package [117] to carry out first-principles calculations on these systems both with and without SOC. The PBE generalized-gradient approximation (GGA) is taken to treat the exchange-correlation functional [118, 119], and norm-conserving pseudopotentials are generated from OPIUM package [121, 122]. The energy cutoff is taken as 65 Ry for In_2Se_3 and 55 Ry for Bi_2Se_3 and Sb_2Se_3, with an 8×8×8 Monkhorst-Pack **k** mesh [123]. The wavefunctions defined in the plane-wave basis are extracted from these calculations and Eq. (5.6) is applied to evaluate the spillage.

As mentioned in Sec. 5.1, the spillage can also be calculated in the Wannier basis. Starting from the first-principles calculations, we use the WANNIER90 package [124] to

construct Wannier functions (WFs) and a corresponding realistic TB model [2] for each of the three materials. The basis WFs are constructed by projecting 30 atomic p trial orbitals onto the Bloch subspace of p-like bands to generate a 30-band spinor model for Bi_2Se_3 and Sb_2Se_3, whereas four additional In $5s$ projectors and bands are included in the model for In_2Se_3. In order that they will retain their atomic-like identity as much as possible, the projected WFs are not optimized to minimize the spread functional [9]. We find that the WFs generated by this projection method are almost the same for the systems with and without SOC, so that the matrix elements $M_{mn}(\mathbf{k})$ defined in Sec. 5.2.1 can be evaluated with good accuracy.

The spillage from the direct plane-wave calculations are shown as the solid lines in Fig. 5.5(a). For Bi_2Se_3, the spillage $\gamma(\mathbf{k})$ has a peak value of 2.12 at Γ, which is slightly larger than 2, indicating that two Kramers degenerate bands at Γ have been inverted by SOC. On the other hand, the effect of SOC in In_2Se_3 and Sb_2Se_3 seems to be negligible everywhere in the BZ, which is consistent with the fact that they are both trivial insulators.

The calculations carried out in the Wannier basis are shown by the dashed lines in Fig. 5.5(a). The spillage is typically slightly larger for the direct plane-wave calculations, since the fact that the WFs have a slightly different plane-wave representation with and without SOC is not taken into account in the Wannier-based calculations. Still, the qualitative features are the same, showing that the Wannier-based approach can successfully provide the same kind of information about the nature and location of the topological band inversion. In Fig. 5.5(b) we also show the spillage of Bi_2Se_3 in the (k_x, k_y) plane at $k_z = 0$, as calculated in the Wannier basis, which again indicates a highly localized band inversion near Γ and is fully consistent with the expected picture of the band inversion in Bi_2Se_3.

[2]The TB models from Wannier90 are constructed in such a way that the Wannier-interpolated bandstructure reproduces the first-principles bandstructure exactly within an energy window centered around the Fermi level.

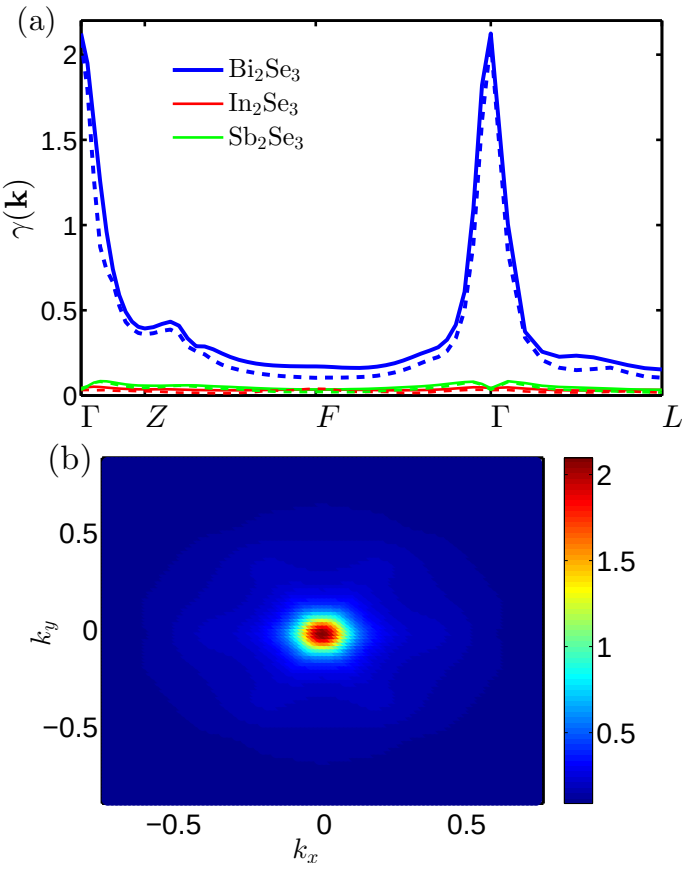

Figure 5.5: (a) Spin-orbit spillage of rhombohedral Bi_2Se_3, Sb_2Se_3 and In_2Se_3 as indicated by blue, green and red lines respectively. Solid (dashed) lines show the spillage computed from direct plane-wave (Wannier-based) calculations. (b) Spillage of Bi_2Se_3 in the (k_x, k_y) plane at $k_z = 0$ (units of Å^{-1}).

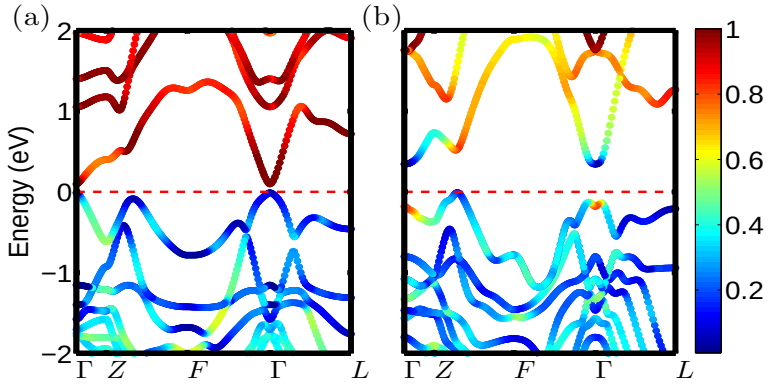

Figure 5.6: (a) Wannier-interpolated bandstructure of Sb_2Se_3. (b) Same for Bi_2Se_3. Color coding indicates weight of Sb $5p$ or Bi $6p$ orbitals.

To see the band inversion from another perspective, we plot in Fig. 5.6 the bulk bandstructures of Sb_2Se_3 and Bi_2Se_3 projected onto Sb $5p$ and Bi $6p$ orbitals respectively. It is clear that for Sb_2Se_3, the Sb $5p$ orbitals are almost all concentrated in the conduction bands, whereas in Bi_2Se_3 there is a localized region around Γ where the corresponding Bi $6p$ orbitals contribute mostly to the top valence band. This is precisely the region of the band inversion corresponding to the peak at Γ in Fig. 5.5.

5.4 Summary

To summarize, we have introduced the SOC-induced spillage $\gamma(\mathbf{k})$ as a useful quantitative tool for evaluating the degree of band inversion driven by SOC and mapping it as a function of $\mathbf{k}$ in the BZ. We have applied this method to the two-band Dirac model, the 2D Kane-Mele model and a tight-binding model of a Bi bilayer with applied Zeeman field, as well as to realistic materials including both trivial and topological insulators. A clear correspondence between non-trivial topological indices and non-trivial spillage distributions is evident. In the two-band Dirac model, one observes interesting behavior in the distribution of spillage through a topological phase transition process. In the Kane-Mele model, one gets two peaks of spillage at K and K$'$ with the peak value of 1, which indicates that a single band is inverted at these two points corresponding to an odd 2D $\mathbb{Z}_2$ index. In the Bi bilayer with applied Zeeman field, a peak of spillage shows up at Γ, with the peak value corresponding to the absolute value of the Chern number. In Bi_2Se_3, the spillage is slightly greater than 2 at one of the TRIM, namely Γ, implying that two bands are inverted by SOC there and signaling the presence of a nontrivial strong $\mathbb{Z}_2$ index.

As mentioned above, other methods exist for the direct computation of topological Chern and $\mathbb{Z}_2$ indices, with or without inversion symmetry [164, 90, 130], and we still recommend these if a direct and definitive determination of the topological indices is needed. However, the present spillage-based approach has the advantage of providing

a BZ map of the strength, position, and degree of localization of the band inversion responsible for the topological character, thus giving valuable physical intuition about the origin of the topological properties of the material in question. In addition, compared with direct methods for topological index calculation, the spillage calculation only requires the evaluation of overlaps between two wavefunctions at the same **k** point, which is easy to implement and numerically very efficient. Therefore, it is our hope that the calculation of SOC spillage will prove to be a widely useful tool that can be applied both for high-throughput screening for topological materials and for obtaining a deeper understanding of the critical features of the bandstructures in known topological materials.

This work was supported by NSF Grant DMR-10-05838. We appreciate valuable discussions with Hongbin Zhang and Huaqing Huang.

Chapter 6

Gauge-discontinuity contribution to the Chern-Simons orbital magnetoelectric coupling

As discussed in Chapter 2, there is a geometric term in the magnetoelectric coupling coefficient known as the "Chern-Simons term", or the "θ term" (Eq. (2.82)-(2.83)). In this chapter, we propose a new method for calculating the Chern-Simons orbital magnetoelectric coupling coefficient. According to previous theories, θ can be expressed as a 3D Brillouin-zone integral of the Chern-Simons 3-form defined in terms of the occupied Bloch functions. Such an expression is valid only if a smooth and periodic gauge has been chosen in the entire Brillouin zone. However, previous calculations show that for systems expected to possess giant Chern-Simons magnetoelectric response, such as topological insulators and systems derived from topological insulators, it is difficult to get numeric convergence for the 3D integral. In order to resolve this problem, we propose to relax the periodicity condition in one $\mathbf{k}$ direction (say, the k_z direction) so that some gauge discontinuity is introduced on a 2D $\mathbf{k}$ plane (normal to k_z). The total θ response is then contributed from both the integral of the Chern-Simons 3-form over the 3D bulk BZ and the gauge discontinuity expressed as a 2D integral over the $\mathbf{k}$ plane. Sometimes the 2D $\mathbf{k}$ plane where the gauge discontinuity resides may be further divided into subregions by 1D "vortex loops", which also contribute to the total θ as Berry phases around the vortex loops. The total θ thus consists of three terms which can be expressed as integrals over 3D, 2D and 1D manifolds. Interestingly, in the presence of time-reversal symmetry, both the 3D and 2D integrals vanish due to a time-reversal invariant gauge choice, and the 1D vortex-loop integral is proved to

be either 0 or π, corresponding to the $\mathbb{Z}_2$ classification of 3D time-reversal invariant insulators. Our method has been successfully applied to the Fu-Kane-Mele model with applied staggered Zeeman field.

6.1 Introduction

As introduced in Sec. 2.2.3, the total magnetoelectric (ME) response is contributed by two terms. The first term stems from lattice degrees of freedom, and the second is from the electronic degrees of freedom with frozen ionic motions. Each of the two contributions can be further divided into spin and orbital components. We are only interested in the frozen-ion orbital ME coupling in this chapter. The frozen-ion orbital ME coupling also consists of two terms. The first term is a conventional response function known as the "Kubo term". The second term is isotropic, and is completely determined by the geometric property of the ground-state wavefunction, which is known as the Chern-Simons term or the θ term [74, 71].

The Chern-Simons orbital ME coupling has drawn significant attention recently due to the interest in topological phases of matter in condensed matter physics [34]. It has been shown that for topological insulators (TIs) with time-reversal ($\mathcal{T}$) symmetry and axion insulators with inversion ($\mathcal{P}$) symmetry [165, 76], both the spin ME response and the Kubo-like orbital ME response have to vanish due to the symmetries. However, there is still an exotic isotropic ME response quantized as $\pm e^2/2h$ in these two phases, which is exactly the Chern-Simons term [166, 20, 76].

As mentioned in Sec. 2.2.3, the Chern-Simons coupling is conventionally scaled by a dimensionless parameter θ via

$$\alpha_{ij}^{\mathrm{CS}} = \frac{\theta e^2}{2\pi h}\delta_{ij} \ . \tag{6.1}$$

θ can be expressed as an integral of the Chern-Simons 3-form over the 3D Brillouin

zone (BZ)

$$\theta = -\frac{1}{4\pi}\int d^3k\ \varepsilon_{abc}\mathrm{Tr}\left[A_a\partial_b A_c - \frac{2}{3}iA_aA_bA_c\right], \tag{6.2}$$

where A_a, A_b and A_c are the Berry connection matrices of the occupied Bloch bands, and the trace is taken over the occupied bands. For TIs and axion insulators, $\theta = \pm\pi$. In the more general cases with the breaking of both $\mathcal{T}$ and $\mathcal{P}$ symmetries, θ is no longer quantized as $\pm\pi$, and other components of ME response would also contribute.

Despite the intriguing properties of the Chern-Simons coupling as discussed in detail in Sec. 2.2.3, up to now it is still a challenging problem to accurately calculate θ using Eq. (6.2) for a specific system. For example, as reported in Ref. [95], the calculated θ for Bi_2Se_3, one of the prototype TIs, on a 11×11×11 first-principles $\mathbf{k}$ mesh is only $\sim 35\%$ of π. In Ref. [20], the authors calculated the ME response of the Fu-Kane-Mele model with applied staggered Zeeman field. As the system approaches to the TI phase, however, the authors switched to some indirect methods to compute θ, because a direct numeric implementation of Eq. (6.2) became difficult to converge. In other words, despite the theoretical importance, Eq. (6.2) is not suitable for practical calculations.

The essential problem is that the integrand in Eq. (6.2) is gauge dependent. As a result, in order to numerically implement Eq. (6.2) on a discrete $\mathbf{k}$ mesh, one has to adopt a smooth and periodic gauge over the entire 3D BZ. On the other hand, as is well known, a nontrivial topological indices usually brings some obstruction against constructing a smooth and periodic gauge in the BZ. For example, for a 2D quantum anomalous Hall (QAH) insulator (such as the Haldane model [24]) with non-zero Chern number, it is just impossible to construct a smooth and periodic gauge in the entire 2D BZ. This implies Eq. (6.2) would completely break down for a 3D analogue of the QAH insulators [1], which is thus beyond the scope of the present work. For 2D and

[1] A 3D analogue of QAH insulators is defined as a 3D insulator whose Chern number of each 2D $\mathbf{k}$ slice in 3D BZ is non-zero and equals to each other. Such a system can be realized by stacking QAH layers in the third spatial dimension, assuming that the inter-layer coupling is weak so that the topologically nontrivial gap remains unclosed.

3D $\mathbb{Z}_2$ TIs, it is impossible to construct a smooth and periodic gauge respecting $\mathcal{T}$ symmetry throughout the BZ [56, 57], although in principle a smooth and periodic gauge breaking $\mathcal{T}$ symmetry is allowed [57]. As a result of the topological obstruction, for $\mathbb{Z}_2$ TIs, or for systems with strong spin-orbit coupling and are close to the $\mathbb{Z}_2$-odd phase, the constriant of being both smooth and periodic is typically too strong such that the gauge would have to be strongly twisted in the BZ to satisfy both conditions. This makes the numeric implementation of Eq. (6.2) difficult.

In this chapter, we propose a new method to compute the Chern-Simons orbital ME coefficient. The general idea is to relax the periodicity condition of the gauge in one direction, say the k_z direction, thus introducing some gauge discontinuity at a 2D $\mathbf{k}$ plane (normal to k_z) denoted by $\mathcal{S}_{\mathrm{GD}}$. Then the total θ is the sum of the bulk-BZ integral (Eq. (6.2)) and an extra contribution from the gauge discontinuity. Furthermore, as will be shown in Sec. 6.4, $\mathcal{S}_{\mathrm{GD}}$ may also be divided into subregions by 1D "vortex loops" (Sec. 6.4.1), which contribute to the total θ as Berry phases around the vortex loops. Then the total θ can be expressed as the sum of the 3D integral over the bulk BZ (θ_{BK}), the 2D integral over the gauge-discontinuity plane (θ_{GD}) and the 1D integral(s) over the vortex loop(s) (θ_{VL}).

The above method may be generalized to the situations where the BZ is divided into multiple subvolumes. These subvolumes meet at multiple 2D surface patches where the gauge discontinuities reside. Furthermore, the 2D surface patches may meet at some 1D curves, defined as vortex lines. Again, the subvolumes, the surface patches and the vortex lines should all contribute to the total θ. However, the definition of a vortex line becomes trickier in this more generalized case, and we leave it for future study.

The advantage of our method is that the gauge can be made smoother in the bulk BZ without worrying too much about the periodicity condition, so that it is much easier to get numeric convergence using Eq. (6.2). The loss of periodicity is then compensated by contributions from the gauge discontinuity expressed as 2D surface integrals and

1D vortex-loop integrals. We will show that the formula for these two terms are very simple, and are efficient for the purpose of numeric implementations.

This chapter is organized as follows. In Sec. 6.2, we review the definitions of Berry connections, Berry curvatures and the bulk formula for θ. We also put the main idea into a more specific context and make a formal statement of the problem. In Sec. 6.3, we derive a formula for θ_{GD}, which is expressed as 2D integral over the boundary where the gauge-discontinuity resides. We also discuss about the properties of the formula. In Sec. 6.4, we give a formal definition of a vortex loop, and derive a formula for the vortex-loop contribution to θ (θ_{VL}). We also show that the π quantization of θ in TIs is purely contributed by the vortex-loop term. In Sec. 6.5, we apply the method to the Fu-Kane-Mele model with a staggered Zeeman field. In Sec. 6.6, we make a summary.

6.2 Preliminaries

In this section, we first review the definitions of some basic quantities such as Berry curvatures and Berry connections which will be frequently used in the chapter. The bulk formula of θ Eq. (6.2) is written in a more explicit form. We also explain the main idea in more detail, and make a formal statement of the problem and the goal.

6.2.1 Definitions

We adopt the following definitions. The Berry connection matrix is

$$A_{a,mn}(\mathbf{k}) = i\langle u_{m\mathbf{k}}|\partial_a|u_{n\mathbf{k}}\rangle \ , \tag{6.3}$$

where m and n run over the occupied bands and a and b run over $\{x, y, z\}$. Here the wavevector components k_x etc. are rescaled to run over $[0, 2\pi]$. We shall start dropping the explicit $\mathbf{k}$ arguments and subscripts, keeping in mind that everything is a function of $\mathbf{k}$. Then the non-covariant Berry curvature tensor is

$$\Omega_{ab,mn} = i\,\langle\partial_a u_m|\partial_b u_n\rangle - i\,\langle\partial_b u_m|\partial_a u_n\rangle \ , \tag{6.4}$$

and the covariant Berry-curvature tensor is

$$\widetilde{\Omega}_{ab,mn} = \Omega_{ab,mn} - i[A_a, A_b]_{mn} \ . \tag{6.5}$$

The Chern-Simons coupling θ has been defined in Eq. (6.2), where the trace is over the occupied band indices. Using the cyclic property of the trace, Eq. (6.2) can be written in a more explicit form

$$\theta = -\frac{1}{4\pi} \int d^3k \ \mathrm{Tr} \left[A_x \Omega_{yz} + A_y \Omega_{zx} + A_z \Omega_{xy} - 2i[A_x, A_y] A_z \right] . \tag{6.6}$$

We can also choose to write one of the Berry curvatures in gauge-covariant form:

$$\theta = -\frac{1}{4\pi} \int d^3k \ \mathrm{Tr} \left[A_x \Omega_{yz} + A_y \Omega_{zx} + A_z \widetilde{\Omega}_{xy} - i[A_x, A_y] A_z \right] . \tag{6.7}$$

Eq. (6.7) turns out to be convenient for the derivation of θ_{GD}, as will be shown in Sec. 6.3.

6.2.2 Statement of the problem

Assume that one particular smooth gauge has been chosen in the region $0 \leq k_z < \pi$ and a different one in the region $-\pi < k_z < 0$. The gauge is constructed in such a way that it smoothly evolves from one to the other at $k_z = 0$, but with some gauge discontinuity left at the $k_z = \pm\pi$ plane. (The location of the boundary can easily be generalized.) From now on $\mathbf{k} = (k_x, k_y)$ denotes a point in the 2D slice at $k_z = \pm\pi$, and $|u^{(0)}\rangle$ and $|u^{(1)}\rangle$ denote the wavefunctions defined using the gauges at the "bottom" ($k_z = -\pi$) and "top" ($k_z = \pi$) planes respectively. The corresponding Berry potentials are $A_x^{(0)}$ and $A_y^{(0)}$ from the bottom plane, and $A_x^{(1)}$ and $A_y^{(1)}$ from the top plane. Moreover, we know the unitary matrix $U(\mathbf{k})$ relating the two gauges:

$$\begin{aligned} |\psi^{(1)}_{m\mathbf{k}}\rangle &= \sum_n |\psi^{(0)}\rangle_{n\mathbf{k}} \, U_{nm}(\mathbf{k}) \\ |u^{(1)}_{m\mathbf{k}}\rangle &= e^{-i2\pi z} \sum_n |u^{(0)}_{n\mathbf{k}}\rangle \, U_{nm}(\mathbf{k}) \ . \end{aligned} \tag{6.8}$$

Our goal is to calculate the contribution θ_{GD} coming from this gauge discontinuity, such that if we add this contribution to the bulk volume integral θ_{BK} as in Eq. (6.6), we get the correct total θ. Later, we shall see that there may also be a contribution θ_{VL} from vortex loops around which the gauge discontinuity circulates by an integer multiple of 2π, so that the total axion coupling is given by

$$\theta = \theta_{\mathrm{BK}} + \theta_{\mathrm{GD}} + \theta_{\mathrm{VL}} \ , \tag{6.9}$$

i.e., a sum of contributions evaluated on 3D, 2D, and 1D manifolds.

6.3 Calculation of θ_{GD} on a planar surface

In this section, we derive a formula for θ_{GD} and discuss various properties of the formula. We assume, as above, that the gauge discontinuity occurs on the $k_z = \pm\pi$ plane as schematically shown in Fig. 6.1, and is described by the unitary matrices $U_{\mathbf{k}}$ as a function of $\mathbf{k}$ lying in the 2D plane. We let

$$U(\mathbf{k}) = e^{-iB(\mathbf{k})} \tag{6.10}$$

where $B(\mathbf{k})$ is a Hermitian matrix that varies smoothly with $\mathbf{k}$ in the 2D plane. Note that $B(\mathbf{k})$ is basically just $i\ln(U(\mathbf{k}))$, but a set of branch choices is involved in picking a particular B. That is, in the representation that diagonalizes B, we can add $2\pi n_j$ to the j'th eigenvalue without changing U (n_j is an arbitrary integer). For now we insist that the branch choice is made in such a way that $B(\mathbf{k})$ is continuous, with no 2π discontinuities in any of its eigenvalues throughout the 2D $\mathbf{k}$ plane, but this condition will be relaxed in Sec. 6.4.

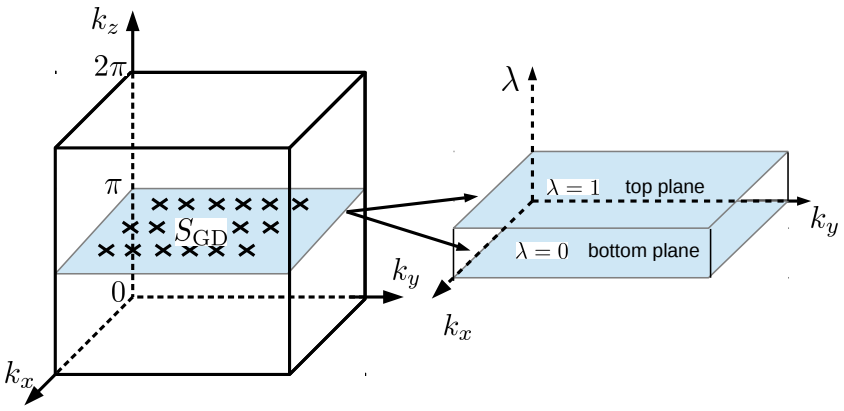

Figure 6.1: A Schematic figure showing that a 2D plane with gauge discontinuity is introduced into the 3D BZ. The two different gauge choices at the top and bottom planes are then smoothly interpolated by introducing a fictitious parameter $\lambda \in [0, 1]$

6.3.1 Formalism

Our strategy is to introduce a parameter λ and define $|u_{m\mathbf{k}}(\lambda)\rangle$ in such a way that it smoothly interpolates from one gauge to the other (Fig. 6.1), i.e.,

$$|u_{m\mathbf{k}}(\lambda)\rangle = e^{-i2\pi z} \sum_n |u^{(0)}_{n\mathbf{k}}\rangle \, W_{nm}(\mathbf{k}, \lambda) \tag{6.11}$$

where

$$W(\mathbf{k}, \lambda) = e^{-i\lambda B(\mathbf{k})} \tag{6.12}$$

so that $W(\mathbf{k}, 0) = I$ and $W(\mathbf{k}, 1) = U(\mathbf{k})$. It is worth noting that $W(\mathbf{k}, \lambda)$ commutes with $B(\mathbf{k})$. We shall again dropping the $\mathbf{k}$ labels, and will frequently use W and B below.

We then calculate the gauge-discontinuity contribution to θ, denoted by $\theta_{\rm GD}$, by integrating Eq. (6.7) over the region $\lambda \in [0, 1]$, where Eq. (6.7) is applied in (k_x, k_y, λ) space instead of (k_x, k_y, k_z) space. A straightforward set of calculations shows that

$$A_x(\lambda) = W^\dagger(\lambda) \, A^{(0)}_x \, W(\lambda) + iW^\dagger(\lambda) \, \partial_x W(\lambda) \, , \tag{6.13}$$

$$A_y(\lambda) = W^\dagger(\lambda) \, A^{(0)}_y \, W(\lambda) + iW^\dagger(\lambda) \, \partial_y W(\lambda) \, , \tag{6.14}$$

$$A_\lambda(\lambda) = B \, . \tag{6.15}$$

where $A^{(0)}_{x(y)}$ is the Berry connection evaluated at the bottom plane as defined earlier.

Then we can write

$$\theta_{\rm GD} = -\frac{1}{4\pi}\int d^2k\; G(\mathbf{k})\;, \tag{6.16}$$

where

$$G = \int_0^1 d\lambda\; {\rm Tr}\left[A_x\Omega_{y\lambda} - A_y\Omega_{x\lambda} + A_\lambda\widetilde{\Omega}_{xy} - i[A_x, A_y]A_\lambda\right]\;. \tag{6.17}$$

We can then write G as the sum of three contributions, $G = G_1 + G_2 + G_3$, where

$$G_1(\mathbf{k}) = \int_0^1 d\lambda\; {\rm Tr}\left[B\,\widetilde{\Omega}_{xy}\right]\;, \tag{6.18}$$

$$G_2(\mathbf{k}) = \int_0^1 d\lambda\; {\rm Tr}\left[A_x\Omega_{y\lambda} - A_y\Omega_{x\lambda}\right]\;, \tag{6.19}$$

$$G_3(\mathbf{k}) = \int_0^1 d\lambda\; {\rm Tr}\Big[-i\left[A_x, A_y\right]B\Big]\;. \tag{6.20}$$

The G_1 term is easy; because $\widetilde{\Omega}_{xy}$ is gauge-covariant, it follows that $\widetilde{\Omega}_{xy}(\lambda) = W^\dagger(\lambda)\,\widetilde{\Omega}^{(0)}_{xy}\,W(\lambda)$ and since $[B, W(\lambda)] = 0$ the integrand is independent of λ, so that

$$G_1(\mathbf{k}) = {\rm Tr}\left[B\,\widetilde{\Omega}^{(0)}_{xy}\right]\;. \tag{6.21}$$

In order to evaluate G_2 and G_3, we need to evaluate objects such as $\partial_x W(\lambda)$ in Eq. (6.13), which can be done by noting that the derivative of an exponential of a matrix can be written as

$$\partial_x e^{-i\lambda M} = -i\int_0^\lambda d\mu\; e^{-i\,\mu M}(\partial_x M)e^{-i\,(\lambda-\mu)M}\;. \tag{6.22}$$

This motivates us to define

$$\overline{B}_j(\lambda) = \int_0^\lambda d\mu\; e^{-i\mu B}\,B_j\,e^{i\mu B}\;, \tag{6.23}$$

so that Eq. (6.12) gives

$$\partial_j W(\lambda) = \partial_j e^{-i\lambda B} = -i\overline{B}_j(\lambda)W(\lambda)\;, \tag{6.24}$$

where $j = x, y$. Then Eqs. (6.13-6.14) become

$$A_j(\lambda)\; = W^\dagger\,\mathcal{A}_j\,W\;, \tag{6.25}$$

where

$$\mathcal{A}_j = A_j^{(0)} + \overline{B}_j \ . \tag{6.26}$$

Again, $j = x, y$, and the dependence on λ is implicit.

Now for the G_2 term we need to compute terms like $\partial_\lambda A_x$. Using Eq. (6.25) and $\partial_\lambda W(\lambda) = -iBW(\lambda)$, it becomes

$$\partial_\lambda A_x \ = iW^\dagger \left[B, \mathcal{A}_x\right] W + B_x \ . \tag{6.27}$$

Recalling that $\Omega_{x\lambda} = \partial_x A_\lambda - \partial_\lambda A_x$ and $\partial_x A_\lambda = B_x$, we get a nice cancellation, and can write

$$\Omega_{x\lambda} = -i\, W^\dagger \left[B, \mathcal{A}_x\right] W \ , \Omega_{y\lambda} = -i\, W^\dagger \left[B, \mathcal{A}_y\right] W \ .$$

Substituting these expressions into Eq. (6.20) then gives

$$G_2(\mathbf{k}) = \int_0^1 d\lambda \ \mathrm{Tr}\left[2i\, B \left[\mathcal{A}_x, \mathcal{A}_y\right]\right] . \tag{6.28}$$

As it happens, this is almost the same as the expression for G_3 in Eq. (6.20). Since B commutes with W we can use the representation-invariance and cyclic properties of the trace to write it as

$$G_3(\mathbf{k}) = \int_0^1 d\lambda \ \mathrm{Tr}\left[- i\, B \left[\mathcal{A}_x, \mathcal{A}_y\right]\right] . \tag{6.29}$$

Thus, this term cancels half of G_2. Restoring the explicit λ dependencies, we get

$$G = \mathrm{Tr}\left[B\left(\widetilde{\Omega}_{xy}^{(0)} + i\int_0^1 d\lambda \left[\mathcal{A}_x(\lambda), \mathcal{A}_y(\lambda)\right]\right)\right] , \tag{6.30}$$

which is a remarkably simple result in the end.

Using Eq. (6.26), this can be written explicitly as

$$G(\mathbf{k}) = \mathrm{Tr}\left[B\Big(\Omega_{xy}^{(0)} + \overline{\overline{B}}_{x,y} + i[\,\overline{\overline{B}}_x, A_y^{(0)}\,] - i[\,\overline{\overline{B}}_y, A_x^{(0)}\,]\Big)\right] , \tag{6.31}$$

where

$$\overline{\overline{B}}_x = \int_0^1 d\lambda \ \overline{B}_x(\lambda) \ , \tag{6.32}$$

$$\overline{\overline{B}}_y = \int_0^1 d\lambda \ \overline{B}_y(\lambda) \ , \tag{6.33}$$

$$\overline{\overline{B}}_{x,y} = i\int_0^1 d\lambda \ [\,\overline{B}_x(\lambda), \overline{B}_y(\lambda)\,] \ . \tag{6.34}$$

Eq. (6.31) is one of our central results of this chapter.

We would like to make some remarks on the formula for θ_{GD}. First of all, it can be analytically proved that the choice of "bottom" and "top" surfaces (Fig. 6.1) is indeed arbitrary (details not shown). In other words, the formula remains invariant under the reverse of the integration direction along λ axis. Secondly, in the case that $B(\mathbf{k})$ can be made diagonal at all $\mathbf{k}$ (i.e., globally diagonal) [2], all the quantities such as W, $B_{x(y)}$ and $\overline{B}_{x(y)}$ commute with each other. Thus the only non-vanishing term in Eq. (6.31) is $G = G_1 = \mathrm{Tr}\left[B\Omega^{(0)}_{xy}\right]$. This is actually also the formula when there is only one occupied band. Lastly, we would like to point out that the integration over the λ axis can be carried out analytically in the basis that $B(\mathbf{k})$ is locally diagonal (details given in Appendix 6.7.2). Therefore only a 2D discrete integration over the $\mathbf{k}$ plane is needed, which is numerically efficient.

In the following subsection, we discuss about the property of the formula for θ_{GD} (Eq. (6.16) and Eq. (6.31)) in the presence of $\mathcal{T}$ symmetry. We show that for systems with $\mathcal{T}$ symmetry, if a TR-symmetric gauge has been chosen in the bulk BZ and assuming that $B(\mathbf{k})$ varies smoothly in the 2D $\mathbf{k}$ plane, both θ_{BK} and θ_{GD} must vanish.

6.3.2 Time-reversal symmetry

Let us consider the situation that the system has $\mathcal{T}$ symmetry, and that a smooth gauge respecting $\mathcal{T}$ symmetry has been chosen in the bulk. For a general $\mathcal{T}$-invariant system with $2N$ occupied bands, one can always decompose the $2N$ bands into N Kramers pairs. A gauge respecting $\mathcal{T}$ symmetry means that for the nth pair ($n \leq N$), the following relationship is satisfied:

$$\begin{aligned} \mathcal{T}|u_{n\mathbf{k},1}\rangle &= -|u_{n-\mathbf{k},2}\rangle \\ \mathcal{T}|u_{n\mathbf{k},2}\rangle &= |u_{n-\mathbf{k},1}\rangle \end{aligned} \tag{6.35}$$

[2] It is unclear, however, how B can be made globally diagonal.

where the indices "1" and "2" are the spin or pseudospin indices in each Kramers pair, and $\mathcal{T}$ denotes the time-reversal operator. If the gauge choice in the bulk satisfies Eq. (6.35), it turns out the Berry curvatures and Berry connections obey the following relationship:

$$\begin{aligned} A_i(\mathbf{k}) &= \sigma_y \left(A_i(-\mathbf{k})\right)^T \sigma_y, \\ \Omega_{ij}(\mathbf{k}) &= -\sigma_y \left(\Omega_{ij}(-\mathbf{k})\right)^T \sigma_y \end{aligned} \tag{6.36}$$

where $i, j = x, y, z$ label the components of $\mathbf{k}$ vectors. The Pauli matrix σ_y acts only on the (pseudo) spin subspace, and the superscript "T" refers to matrix transpose. Since the Berry curvature behaves as an odd function while the Berry connections behave as an even function of $\mathbf{k}$, it is easy to show that both $\mathrm{Tr}\left[A_i(\mathbf{k})\Omega_{ij}(\mathbf{k})\right]$ and $\mathrm{Tr}\left[iA_i(\mathbf{k})\left[A_j(\mathbf{k}), A_k(\mathbf{k})\right]\right]$ cancel their time-reversal partners at $-\mathbf{k}$. Therefore, the bulk integral θ_{BK} (Eq. (6.6)) vanishes if a smooth $\mathcal{T}$-symmetric gauge is chosen in the bulk BZ.

In particular, at the boundary of the BZ where there is gauge discontinuity, the wavefunctions at the bottom and top planes of the BZ are connected via $\mathcal{T}|u^{(0)}_{n\mathbf{k},1}\rangle = -|u^{(1)}_{n-\mathbf{k},2}\rangle$, $\mathcal{T}|u^{(0)}_{n\mathbf{k},2}\rangle = |u^{(1)}_{n-\mathbf{k},1}\rangle$ ($\mathbf{k}$ now is interpreted as wavevectors in 2D plane). With such a $\mathcal{T}$-symmetric gauge choice, the B matrix, the Berry connections, and the Berry curvature satisfy the following relationship:

$$B(\mathbf{k}) = \sigma_y B(-\mathbf{k})^T \sigma_y, \tag{6.37}$$

$$A^{(0)}_x(\mathbf{k}) = \sigma_y \left(A^{(1)}_x(-\mathbf{k})\right)^T \sigma_y, \tag{6.38}$$

$$A^{(0)}_y(\mathbf{k}) = \sigma_y \left(A^{(1)}_y(-\mathbf{k})\right)^T \sigma_y, \tag{6.39}$$

$$\Omega^{(0)}_{xy}(\mathbf{k}) = -\sigma_y \left(\Omega^{(1)}_{xy}(-\mathbf{k})\right)^T \sigma_y \tag{6.40}$$

Again, the superscript "(0)" and "(1)" refer to the quantities evaluated at $\lambda = 0$ and $\lambda = 1$ respectively. We claim that if Eq. (6.37)-(6.40) are satisfied, and if all the quantities involved in the formula vary smoothly in the 2D BZ, then θ_{GD} must vanish, which will be proved as follows.

First of all, it is easy to see that the first term in Eq. (6.30) vanishes due to $\mathcal{T}$ symmetry. Note that the gauge covariant Berry curvature $\widetilde{\Omega}^{(1)}_{xy} = U^\dagger\, \widetilde{\Omega}^{(0)}_{xy}\, U$ and $U = e^{-iB}$ commutes with B, so $\mathrm{Tr}\left[B\widetilde{\Omega}^{(1)}_{xy}\right] = \mathrm{Tr}\left[B\widetilde{\Omega}^{(0)}_{xy}\right]$. On the other hand, B transforms as an even function of $\mathbf{k}$ while $\widetilde{\Omega}_{xy}$ transforms as an odd function of $\mathbf{k}$. This leads to the cancellation at $\mathbf{k}$ and $-\mathbf{k}$ for the G_1 term.

The second term in Eq. (6.30) is trickier. First, from the representation-invariance of the trace and the fact $W = e^{-i\lambda B}$ commutes with B, we know that $\mathrm{Tr}\left[i\, B\, [\mathcal{A}_x, \mathcal{A}_y]\right] = \mathrm{Tr}\left[i\, B\, [A^{(\lambda)}_x, A^{(\lambda)}_y]\right]$. Then we claim that the Berry connection matrix at (k_x, k_y, λ) is connected to that at $(-k_x, -k_y, 1-\lambda)$ by a $\mathcal{T}$ transformation

$$A^{(\lambda)}_j(\mathbf{k}) = \sigma_y \left(A^{(1-\lambda)}_j(-\mathbf{k})\right)^T \sigma_y \ , \tag{6.41}$$

where $A^{(\lambda)}_j \equiv A_j(\lambda)$, with $j = x, y$, as defined in Eq. (6.13)-(6.14). Eq. (6.41) will be proved in Appendix 6.7.3, but if one considers λ as the third wavevector component, Eq. (6.41) is indeed very intuitive. Combing Eq. (6.41) and Eq. (6.25), it follows that:

$$\eta(\mathbf{k}, \lambda) = -\eta(-\mathbf{k}, 1-\lambda) \ , \tag{6.42}$$

where

$$\eta(\mathbf{k}, \lambda) = \mathrm{Tr}\left[\, i\, B(\mathbf{k})\, [\mathcal{A}^{(\lambda)}_x(\mathbf{k}), \mathcal{A}^{(\lambda)}_y(\mathbf{k})]\,\right] \ . \tag{6.43}$$

Here $\mathcal{A}^{(\lambda)}_{x(y)} \equiv \mathcal{A}_{x(y)}(\lambda)$. Therefore, the integrand at $(\mathbf{k}, \lambda)$ and $(-\mathbf{k}, 1-\lambda)$ cancel each other, so that the second term in Eq. (6.30) also vanishes. It is therefore concluded that if $B_{\mathbf{k}}$ varies continuously, both θ_{BK} and θ_{GD} vanish with a $\mathcal{T}$-symmetric gauge choice.

To summarize this section, we have derived a formula for the gauge discontinuity contribution θ_{GD}. The key result is shown in Eq. (6.16) and Eq. (6.30)-(6.31). We also show that for a system with $\mathcal{T}$ symmetry, if a $\mathcal{T}$-symmetric gauge choice is constructed in the bulk, and if the branch choice is made in such a way that $B(\mathbf{k})$ varies smoothly over the entire 2D $\mathbf{k}$ plane, then both θ_{BK} and θ_{GD} must vanish. However, it is well known that $\theta = \pi$ for $\mathbb{Z}_2$ TIs, then where is the quantized θ from? The answer is that in

the $\mathbb{Z}_2$-odd case, it is impossible to insist a branch choice such that B remains smooth throughout the (k_x, k_y) plane. In other words, the 2D $\mathbf{k}$ plane have to be divided into some subregions so that the eigenvalues of B would be shifted by integer multiples of 2π when going from one subregion to another. The boundary of such 2D subregions are called vortex loops. It turns out that for a $\mathbb{Z}_2$-odd TI, the contribution from such a vortex loop is exactly π. We will elaborately discuss about the vortex-loop contribution in the following section.

6.4 Vortex-loop contribution

In this section, we first propose a formal definition for a vortex loop in Sec. 6.4.1. In Sec. 6.4.2, we derive a formula for the vortex-loop contribution θ_{VL}, which turns out to be very simple. It only involves in the Berry phases of the electronic Bloch functions and the "Berry phases" of the eigenvectors of $B(\mathbf{k})$ around the vortex loops. In Sec. 6.4.3 we discuss properties of the formula for θ_{VL}. In particular, we show that in systems with $\mathcal{T}$ symmetry, θ_{VL} must be either 0 or π, corresponding to the $\mathbb{Z}_2$ classification of 3D insulators with $\mathcal{T}$ symmetry.

6.4.1 What is a vortex loop

In Sec. 6.2.2 we suggested that the complete formula for θ should include three kinds of contributions, as expressed by Eq. (6.9). Here we review the philosophy of the calculation, and explaining why the third vortex-loop contribution θ_{VL} may be needed.

First, we choose a smooth gauge in the 3D bulk BZ, but the periodicity condition in the k_z direction is relaxed. Hence some gauge discontinuity is introduced at some 2D boundary plane normal to k_z. The 3D bulk integral (Eq. (6.2)) excluding the boundary is the θ_{BK} term in Eq. (6.9).

Next, we identify the 2D boundary as $\mathcal{S}_{\mathrm{GD}}$. Let us define $\mathcal{S}_{\mathrm{GD}}$ as a directed area with surface normal $\hat{\mathbf{n}}$. In order to compute the integral over the 2D plane $\mathcal{S}_{\mathrm{GD}}$, $\hat{\mathbf{n}}$ is chosen

in such a way that $\hat{\mathbf{x}}$-$\hat{\mathbf{y}}$-$\hat{\mathbf{n}}$ form a right-handed coordinate triad. The gauge discontinuity in the $\hat{\mathbf{n}}$ direction is given by a unitary matrix $U(\mathbf{k}) = e^{-iB(\mathbf{k})}$ which varies smoothly with $\mathbf{k}$ lying in the 2D plane. Since the Hermitian matrix $B(\mathbf{k}) = i\ln(U(\mathbf{k}))$ is involved in the formula of $\theta_{\rm GD}$ (Eq. (6.31)), a branch choice for B has to be made. If possible we make a branch choice so that $B(\mathbf{k})$ is smooth and continuous over the entire $\mathbf{k}$ plane, but this may not always be possible or desirable. In that case $\mathcal{S}_{\rm GD}$ is divided into subregions within each of which $B(\mathbf{k})$ is smooth and continuous. The 2D contribution $\theta_{\rm GD}$ is then computed by integrating over each subregion using Eq. (6.30)-(6.31) of Sec. 6.3.

Since the eigenvalues of B may exhibit abrupt 2π jumps as they vary from one subregion to another, the behavior at the boundaries of the subregions, i.e., at the vortex loops, is singular. Such vortex-loop contributions cannot be computed by the formula for $\theta_{\rm GD}$, hence a new formula is needed to account for it.

Before deriving the formula for $\theta_{\rm VL}$, let us first propose a formal definition for a vortex loop. As shown in Fig. 6.2, consider the case that a 2D plane is divided into two subregions by a loop denoted as $\mathcal{S}_{\rm GD}$ and $\overline{\mathcal{S}}_{\rm GD}$ respectively. We establish a convention for the direction of integration along the loop by picking a unit vector $\hat{\mathbf{t}}$ tangent to the loop as shown by the red arrow in Fig. 6.2 (a), and denote the directions of the two surface normals in order of counter-clockwise circulation relative to $\hat{\mathbf{t}}$, as indicated by the black arrows $\hat{\mathbf{n}}_a$ and $\hat{\mathbf{n}}_b$ in Fig. 6.2.

Let us take an arbitrary point $\mathbf{k}_0$ on the loop whose tangential vector is $\hat{\mathbf{t}}(\mathbf{k}_0)$. Then we take two $\mathbf{k}$ points lying on two different sides of, but infinitely close to $\mathbf{k}_0$, denoted as $\mathbf{k}_0^+$ and $\mathbf{k}_0^-$ respectively. One can define two U matrices $U_a(\mathbf{k}_0^+)$ and $U_b(\mathbf{k}_0^-)$. On one side of the loop, the gauge discontinuity is described by $U_{a,mn}(\mathbf{k}_0^+) = \langle u_m^{(0)} | u_n^{(1)} \rangle = (\, e^{-iB_a(\mathbf{k}_0^+)}\,)_{mn}$; while on the other side of the loop, following the circulation direction around $\hat{\mathbf{t}}(\mathbf{k}_0)$, the unitary matrix is $U_b(\mathbf{k}_0^-) = e^{-iB_b(\mathbf{k}_0^-)} = U_a^\dagger(\mathbf{k}_0^+)$. Obviously, $\Delta U = U_a(\mathbf{k}_0^+)\, U_b(\mathbf{k}_0^-) = 1$. Now we are ready to ask if this loop is a vortex loop. We make a

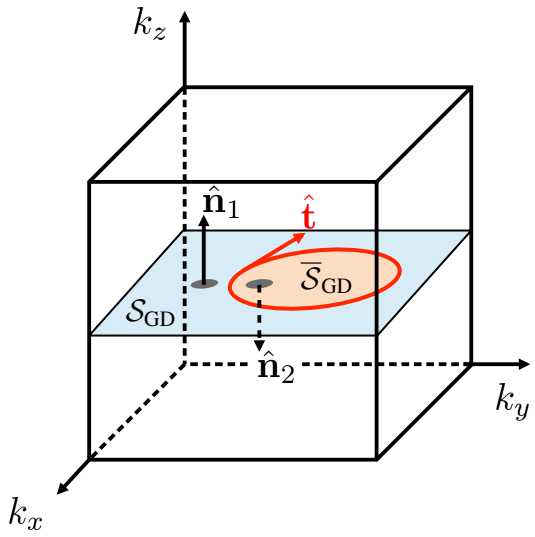

Figure 6.2: A schematic figure showing that there is a 2D $\mathbf{k}$ plane with gauge discontinuity in the 3D BZ. Further more, the 2D plane is divided into two patches $\mathcal{S}_{\mathrm{GD}}$ and $\overline{\mathcal{S}}_{\mathrm{GD}}$ by a vortex loop (red line). The red arrow represents a vector that is tangential to the vortex loop. The two black arrows $\hat{\mathbf{n}}_a$ and $\hat{\mathbf{n}}_b$ indicate the surface normals of $\mathcal{S}_{\mathrm{GD}}$ and $\overline{\mathcal{S}}_{\mathrm{GD}}$ respectively. (a) A 3D view, and (b) a side view

branch choice for B and define the quantity $\Delta B(\mathbf{k}_0) = B_a(\mathbf{k}_0^+) + B_b(\mathbf{k}_0^-) = i \ln U_a(\mathbf{k}_0^+) + i \ln U_b(\mathbf{k}_0^-)$. Since $\mathbf{k}_0^+$ and $\mathbf{k}_0^-$ are infinitely close to each other, $U_a(\mathbf{k}_0^+)$ and $U_b(\mathbf{k}_0^-)$ share the same eigenvectors, thus commute with each other. This means $e^{-i\Delta B(\mathbf{k}_0)} = e^{-i(B_a(\mathbf{k}_0^+)+B_b(\mathbf{k}_0^-))} = e^{-iB_a(\mathbf{k}_0^+)} e^{-iB_b(\mathbf{k}_0^-)} = 1$. Therefore, the eigenvalues of $\Delta B(\mathbf{k}_0)$ can only be an integer multiples of 2π. Then the loop is defined as a vortex loop if at least one eigenvalue of ΔB is non-zero.

We would like to point out that the above definition applies only to the simple case where there is one gauge-discontinuity plane in the BZ, and the vortex loops show up as boundaries of the 2D subregions in the gauge-discontinuity plane. In more general cases, a 3D BZ may be divided into multiple subvolumes, and these subvolumes meet at multiple 2D surface patches with gauge discontinuities. These surface patches may further meet at one or more vortex lines. Now the definition of a vortex line needs to be generalized to a situation that the U matrices from different surface patches do not commute with each other. We leave this more complicated situation to future study.

Certainly the presence or absence of vortex loops is dependent on branch choice.

However, when the system is topologically nontrivial, the presence of vortex loops may also be due to topological obstructions, and becomes independent of branch choice. To be specific, if the system has neither $\mathcal{T}$ nor $\mathcal{P}$ symmetry (assuming vanishing Chern number), then one can generally make a proper branch choice such that all the eigenvalues of B remain smooth throughout the 2D $\mathbf{k}$ plane. This is because in general the eigenvalues of B are non-degenerate everywhere in the absence of $\mathcal{T}$ and $\mathcal{P}$ symmetries.

However, if either $\mathcal{T}$ or $\mathcal{P}$ symmetry is present, there must be degeneracy between different eigenvalues of B at the four time-reversal invariant momenta (TRIM) in the 2D $\mathbf{k}$ plane [3]. As a result, the topological property of the bulk Hamiltonian becomes closely related to the number of vortex loops in the (k_x, k_y) plane. In the same vein as the $\mathbb{Z}_2$ classification based on the number of surface Dirac cones [30], when there is an odd number of vortex loops centered at an odd number of TRIM in the 2D $\mathbf{k}$ plane, the system is $\mathbb{Z}_2$-odd, corresponding to a topological insulator for the case of $\mathcal{T}$ symmetry, and an axion insulator for the case of $\mathcal{P}$ symmetry [76]. Otherwise when the number of vortex loops is even, the system is topologically trivial. In the topologically nontrivial cases (topological insulators and axion insulators), it is impossible to insist on the smoothness for all the eigenvalues of B throughout the 2D $\mathbf{k}$ plane without jumping to different branches. In principle the vortex loop can be made infinitesimally small such that it shrunk to one of the TRIM, but it can never be removed due to the symmetry-protected degeneracy at the TRIM. Therefore, we must consider the contribution from the vortex loops in such topologically nontrivial phases.

On the other hand, the vortex loops may be present even in the topologically trivial cases unless one makes a proper branch choice to remove them. In realistic calculations, however, one usually adopts some default branch choice (e.g., from -π to π), which is not necessarily the one that makes B globally smooth. Then one has to consider both

[3] Let us consider the gauge-discontinuity plane as an isolated 2D BZ without worrying about its k_z value.

θ_{GD} and θ_{VL}. In this regard, it would be nice to have a formula for the vortex loop contribution, so that one can evaluate the gauge-discontinuity contribution to θ at arbitrary branch choices.

In the remainder of this section, we will derive a formula for θ_{VL}, and discuss properties of the formula. We will show that in the presence of time-reversal symmetry, our formula gives $\theta_{\mathrm{VL}} = \pm\pi$ in the $\mathbb{Z}_2$-odd case.

6.4.2 The formula for $\theta_{\mathbf{VL}}$

Let us first consider the topologically trivial case so that we can always find a proper branch choice in which B remains smooth throughout the 2D plane. Assume this has been done, now shift the nth eigenvalue of B by $2\pi\nu(n)$ within region $\mathcal{S}$, thus creating an artificial vortex loop $\mathcal{C}$ whose interior is $\mathcal{S}$. The above operation is equivalent to making a different branch choice. However, a physical quantity should not depend on the branch choice of B, so θ should remain invariant after such an operation. Let us denote the θ as θ_0 before the operation; after the operation, $\theta = \theta_0 + \theta_{\mathrm{VL}} + \theta_{\mathrm{shift}} = \theta_0$, where θ_{VL} is the contribution from the vortex loop which we do not yet know how to calculate, while θ_{shift} is the contribution from shifting the eigenvalues of B within the subregion $\mathcal{S}$. Therefore, the vortex loop contribution θ_{VL} must be equal (in magnitude) and opposite (in sign) to θ_{shift}:

$$\theta_{\mathrm{VL}} = -\theta_{\mathrm{shift}} \ . \tag{6.44}$$

We begin by considering a simple case in which only one of the eigenvalues of B is shifted by 2π within $\mathcal{S}$. We decompose B into a smooth part B_0 and a extra contribution from the 2π shift ΔB, $B = B_0 + \Delta B$. We choose to connect the states at the bottom and top planes in two steps. In the first step,

$$|u^{(\lambda)}\rangle = e^{-i\lambda\Delta B}|u^{(0)}\rangle, \quad \lambda \in [0,1). \tag{6.45}$$

In the second step,

$$|u^{(\lambda)}\rangle = e^{-i(\lambda-1)B_0}|u^{(1)}\rangle, \quad \lambda \in [1,2]. \tag{6.46}$$

Note that now the states at the top plane is $|u^{(2)}\rangle$ instead of $|u^{(1)}\rangle$. In the second step, B_0 is smooth over the entire 2D BZ. One can define $\lambda' = \lambda - 1$, with $\lambda' \in [0,1]$, then the formula derived in Sec. 6.3 applies.

We assume without loss of generality that the first eigenvalue of B b_1 jumps by 2π in the subregion $\mathcal{S}$. Then ΔB can be written as

$$\Delta B = \begin{cases} V\,\Delta_1\,V^\dagger & \text{if } \mathbf{k} \in \mathcal{S} \\ 0 & \text{otherwise} \end{cases} \tag{6.47}$$

where Δ_1 is a $N \times N$ matrix (N is the number of occupied bands), with $(\Delta_1)_{11} = 2\pi$ and all the remaining matrix elements vanishing. $V = (|v_1\rangle, |v_2\rangle, ..., |v_N\rangle)$ is the eigenvector matrix of B_0, with each column being one of the eigenvectors. Plug the expression of ΔB (Eq. (6.47)) into the expression of G (Eq. (6.31)), one would obtain a formula for θ_{shift}, and θ_{VL} is simply the opposite of θ_{shift}. It turns out that many complicated terms cancel each other, and eventually one obtains a surprisingly simple formula

$$\theta_{\text{VL}} = -\theta_{\text{shift}} = \Big(\,\phi_1(\mathcal{C}) + \xi_1(\mathcal{C})\,\Big)/2\,, \tag{6.48}$$

where $\phi_1(\mathcal{C})$ is the Berry phase of $|\overline{u}^{(0)}_{1\mathbf{k}}\rangle$ around the vortex loop $\mathcal{C}$. $|\overline{u}^{(0)}_1\rangle = \sum_{m=1}^{N} |u^{(0)}_{m\mathbf{k}}\rangle\, V_{m1}$ is the first eigenstate which is unitarily transformed by V from the original gauge at the bottom plane ($\lambda = 0$). $\xi_1(\mathcal{C})$ in the above equation is the "Berry phase" of $|v_1\rangle$ around the vortex loop $\mathcal{C}$, with $|v_1\rangle$ being the first column vector of V. In other words,

$$\xi_1 = \iint\limits_{\mathcal{S}} \Lambda_{xy,11} = \oint_{\mathcal{C}} \mathbf{C}_{11} \cdot d\mathbf{k} \tag{6.49}$$

where $\mathbf{C} = (C_x, C_y)$ can be considered as "Berry connections" in the gauge space,

$$C_{x(y)} = iV^\dagger \partial_{x(y)} V\,. \tag{6.50}$$

Λ_{xy} is then the corresponding "Berry curvature",

$$\Lambda_{xy} = \partial_x C_y - \partial_y C_x \ . \tag{6.51}$$

The details of deriving Eq. (6.48) is given in Appendix 6.7.4.

In general, there may be multiple vortex loops $\{\mathcal{C}_i, i = 1, ..L\}$ in the 2D $\mathbf{k}$ plane, and inside the ith vortex loop the nth eigenvalue of B may be shifted by $2\pi\nu_n(i)$ ($\nu_n(i)$ is an integer). Then Eq. (6.48) can be generalized in a straightforward manner

$$\theta_{\mathrm{VL}} = \sum_i \sum_n \Big(\phi_n(\mathcal{C}_i) + \xi_n(\mathcal{C}_i)\Big)\nu_n(i)/2 \tag{6.52}$$

where $\phi_n(\mathcal{C}_i)$ and $\xi_n(\mathcal{C}_i)$ are the two Berry phases of the nth Bloch states and the nth eigenvector of B around the vortex loop $\mathcal{C}_i$. Eq. (6.48), together with its generalized form Eq. (6.52), is the other central result of this chapter.

In the above discussion, we only consider the situation when $\mathbf{k}$ is in the interior of the vortex loop. If $\mathbf{k}$ lies exactly on the vortex loop, the partial derivatives of B behaves as a δ function due to the abrupt 2π jump in the eigenvalues of B across the vortex loop $\mathcal{C}$. Therefore,

$$(\, V^\dagger \, B_{x(y)} \, V \,)_{mn} = 2\pi\nu_n \delta(\mathcal{C})\delta_{mn} \tag{6.53}$$

However, because $\mathbf{k}$ lies exactly on the vortex loop, the eigenvector matrix V does not vary. As a result $B_{x(y)}$ and $\overline{B}_{x(y)}$ both commute with B, and $\overline{B}_x$ also commutes with $\overline{B}_y$. Therefore, there is no contribution from these δ functions exactly on the vortex loop. It is confirmed that Eq. (6.52) is the final result.

6.4.3 Discussions

We continue discussing the properties of Eq. (6.52) in this subsection. We first show that for a given branch choice, Eq. (6.52) is gauge invariant modulo 2π, which is consistent with the 2π ambiguity of θ. Secondly we prove that Eq. (6.52) remains unchanged by interchanging the two steps (Eq. (6.45) and Eq. (6.46)). Lastly we discuss the situation with $\mathcal{T}$ symmetry, and come to the conclusion that $\theta_{\mathrm{VL}} = \pm\pi$ in the $\mathbb{Z}_2$-odd case.

Gauge invariance

We would like to show that at a fixed branch choice, Eq. (6.48) and Eq. (6.52) are indeed gauge invariant modulo 2π. First note that Eq. (6.52) involves the sum of the Berry phases of $|\overline{u}_n^{(0)}\rangle$ and $|v_n\rangle$, so the only gauge freedom is a $U(1)$ gauge transformation acted on $|v_n\rangle$. Let us apply a $U(1)$ gauge transformation to $|v_n\rangle$ such that $|v_n\rangle \to |v_n\rangle e^{i\beta(\mathbf{k})}$. Then since $|\overline{u}_n^{(0)}\rangle = \sum_{m=1}^{N} |u_m^{(0)}\rangle\, V_{mn}$, the same gauge transformation is also applied to $|\overline{u}_n^{(0)}\rangle$: $|\overline{u}_n^{(0)}\rangle \to |\overline{u}_n^{(0)}\rangle e^{i\beta(\mathbf{k})}$ (Note that $|v_n\rangle$ is a column vector whose mth element is V_{mn}). Therefore, if the gauge transformation has a non-zero winding number (denoted by J) such that $\xi_n \to \xi_n + 2\pi J$, ϕ_n has to be changed by $2\pi J$ either. So it is concluded that Eq. (6.52) is gauge invariant upto 2π.

Order of the two steps

In Sec.6.4.2, we decompose B into two parts $B = B_0 + \Delta B$, where B_0 is the smooth part and ΔB is the contribution from the 2π shift which is equal and opposite to the vortex-loop contribution. Then B is treated in two steps in the fictitious λ space. The first step is to deal with ΔB (Eq. (6.45)), and the second step is to treat the smooth part B_0 (Eq. (6.46)). Here we would like to show that Eq. (6.48) and Eq. (6.52) remain correct even if the order of the two steps is reversed.

If the order is reversed, namely, if $|u^{(\lambda)}\rangle = |u^{(0)}\rangle e^{-i\lambda B_0}$ for $\lambda \in [0, 1)$ and $|u^{(\lambda)}\rangle = e^{-i(\lambda-1)\Delta B}|u^{(1)}\rangle$ for $\lambda \in [1, 2]$, then Eq. (6.48) remains unchanged except that the first term ϕ_1 is interpreted as the Berry phase of $|\overline{u}_1^{(1)}\rangle$, where $|\overline{u}^{(1)}\rangle = |u^{(1)}\rangle\, V = |u^{(0)}\rangle\, e^{-iB_0}\, V$. The Berry phases of $|\overline{u}_1^{(1)}\rangle$ and $|\overline{u}_1^{(0)}\rangle$ around the vortex loop $\mathcal{C}$ are indeed exactly the same. This is because the overlap matrix between $|\overline{u}^{(1)}\rangle$ and $|\overline{u}^{(0)}\rangle$: $\langle\overline{u}^{(0)}|\overline{u}^{(1)}\rangle = V^\dagger\, e^{-iB_0}\, V$, is diagonal, whose nth diagonal element is e^{-ib_n}. It means $|\overline{u}_n^{(1)}\rangle = |\overline{u}_n^{(0)}\rangle e^{-ib_n}$. Since b_n is smooth and single-valued everywhere in the 2D BZ, the Berry phase would not change under such a single-band gauge transformation. Therefore, Eq. (6.48) and Eq. (6.52) remain valid even if the order of the two steps (Eq. (6.45)

and Eq. (6.46)) is reversed.

Time-reversal symmetry

We would like to prove that $\theta_{\rm VL}$ must be either $\pm\pi$ or 0 in the presence of $\mathcal{T}$ symmetry. Let us consider the simple case that there is only one vortex loop $\mathcal{C}$ in the 2D $\mathbf{k}$ plane, and that only one of the eigenvalues of B (labelled as the first eigenvalue) is shifted by 2π in the interior of the vortex loop. Suppose that a smooth gauge respecting $\mathcal{T}$ symmetry has been constructed in the bulk so that both bulk integral $\theta_{\rm BK}$ and surface integral $\theta_{\rm GD}$ (Eq. 6.9) vanish, as discussed in Sec. 6.3.2. Due to the $\mathcal{T}$-symmetric gauge (Eq. (6.35)), the B matrix must satisfy Eq. (6.37), with two eigenvalues being degenerate at the four TRIM, i.e., $(0,0)$, $(0,\pi)$, $(\pi,0)$ and (π,π). As a result, the vortex loop $\mathcal{C}$ has to be a $\mathcal{T}$-symmetric loop (Fig. 2.1(b)), which means for any $\mathbf{k}$ on the loop $\mathcal{C}$, $-\mathbf{k}$ is also on the loop. Then it is well known that the Berry phase around such a $\mathcal{T}$-symmetric loop enclosing a degeneracy point is $\pm\pi$, as has been demonstrated in the surface states of TIs and in systems with Rashba spin-orbit splitting. Therefore, it is concluded that $\xi_1 = \pm\pi$ in Eq. (6.48)

It can also be shown that ϕ_1 (Eq. (6.48)) is exactly the same as ξ_1 as a result of $\mathcal{T}$ symmetry. Let us make a branch choice such that the vortex loop is negligibly small. Then the Berry connection of $|\overline{u}_1^{(0)}\rangle$ can be expressed as

$$\begin{aligned}
\overline{A}^{(0)}_{x,11} &= i\,\langle \overline{u}_1^{(0)} | \partial_x \overline{u}_1^{(0)} \rangle \\
&= i \sum_{j,j'=1}^{N} V^*_{j1} \langle u_j^{(0)} | \partial_x u_{j'}^{(0)} \rangle\, V_{j'1} + i \sum_{j=1}^{N} V^*_{j1} \partial_x V_{j1} \\
&= (V^\dagger\, A_x^{(0)}\, V)_{11} + C_{x,11}
\end{aligned} \tag{6.54}$$

where N is the number of occupied bands, $A_x^{(0)}$ is the Berry-connection matrix in the bottom-plane gauge, and C_x is the "Berry connection" in the gauge space as defined in Eq. (6.50). Assume the vortex loop is so small such that the variation of $|u_1^{(0)}\rangle$ within the vortex loop is negligible, then if follows that $A_x^{(0)} = 0$, which means $\overline{A}_{x,11}$ is purely

contributed by the gauge twist, i.e., $\overline{A}^{(0)}_{x,11} = C_{x,11}$. It follows that $\xi_1 = \phi_1 = \pm\pi$ for such a branch choice, and $\theta_{\mathrm{VL}} = \pm\pi$ according to Eq. (6.48).

The above argument remains valid even if the size of the vortex loop increases so that the variation of $|u^{(0)}_1\rangle$ is no longer negligible. First, it should noted that the total response $\theta = \theta_{\mathrm{BK}} + \theta_{\mathrm{GD}} + \theta_{\mathrm{BK}}$ is independent of the branch choice for B. On the other hand, θ_{BK} and θ_{GD} identically vanish for a $\mathcal{T}$-symmetric gauge (Eq. (6.35)), therefore θ_{VL} must remain invariant regardless of the branch choice. In other words, given a $\mathcal{T}$-symmetric gauge in the bulk BZ, θ_{VL} must be quantized as $\pm\pi$ in the $\mathbb{Z}_2$-odd case regardless of the size of the vortex loop, which means the total θ must be quantized as $\pm\pi$.

The above statement remains valid for other gauge choices (rather than the $\mathcal{T}$-symmetric gauge), because the total θ is gauge-invariant modulo 2π. It is thus concluded that for systems with $\mathcal{T}$ symmetry, when there is one vortex loop, our formula Eq. (6.52) predicts that θ must be $\pm\pi$.

The above discussion can be easily generalized to a more general case with multiple vortex loops. Obviously when there is an odd number of vortex loops with an odd number of 2π shifts in the eigenvalues of B, θ is still quantized $\pm\pi$. If there is an even number of vortex loops, they can either enclose an even number of TRIM or fall into $\mathcal{T}$ partners without enclosing any TRIM. In both cases, $\theta = 0$ modulo 2π.

6.5 Applications

In this section, we will apply our method to a specific physical model. We take the Fu-Kane-Mele (FKM) model, which is a 4-band tight-binding model of s electrons on a 3D diamond lattice with the spin-orbit coupling (SOC) showing up as the second-neighbor hopping terms [30]. The model Hamiltonian is

$$H = \sum_{\langle i,j\rangle} t_{ij} c_i^\dagger c_j + i8\lambda_{\mathrm{SO}} \sum_{\langle\langle i,j\rangle\rangle} c_i^\dagger \, \mathbf{s} \cdot (\mathbf{d}^1_{ij} \times \mathbf{d}^2_{ij}) \, c_j \ , \tag{6.55}$$

where t_{ij} is the first-neighbor spin-independent hopping, $\lambda_{\rm SO}$ is the strength of spin-dependent second-neighbor hopping generated by SOC. $\mathbf{d}_{ij}^1$ and $\mathbf{d}_{ij}^2$ are the two first-neighbor bond vectors connecting two second-neighbor sites i and j. $\mathbf{s} = (s_x, s_y, s_z)$ are Pauli matrices representing the electronic spin. In the remaining of the paper, we only consider the case of half filling, i.e., two occupied bands. Setting $t_{ij} = t_0 = 1$ and $\lambda_{\rm SO} = 0.125$, when the lattice symmetry is preserved, the system is a semimetal with gap closures at the three equivalent X points in the BZ . An energy gap would be opened up if a perturbation lowering the lattice symmetry is added, which could be a topologically nontrivial gap. For example, when the first-neighbor hopping along the [111] direction differs from the other three first-neighbor bonds, the system could be either a trivial insulator or a topological insulator depending on the strength of the bond distortion.

In order to test our formalism, we need to consider the general case without $\mathcal{T}$ symmetry. Following Ref. [20], we apply a staggered Zeeman field with opposite signs to the A and B sublattices of the diamond lattice. Moreover, the [111] first-neighbor bond is distorted by changing the corresponding hopping amplitude from t_0 to $3t_0 + \delta$. Both the Zeeman field strength h and the bond distortion strength δ are dependent on a single scaling parameter β: $h = m \sin\beta$, $\delta = m \cos\beta$. Now the Hamiltonian becomes

$$\begin{aligned} H(\beta) \quad &= \sum_{\langle i,j\rangle=[111]} (3t_0 + m\cos\beta)\, c_i^\dagger c_j + \sum_{\langle i,j\rangle\neq[111]} t_0\, c_i^\dagger c_j + \\ &+ i8\lambda_{\rm SO} \sum_{\langle\langle i,j\rangle\rangle} c_i^\dagger\, \mathbf{s}\cdot(\mathbf{d}_{ij}^1 \times \mathbf{d}_{ij}^2)\, c_j + m\sin\beta \sum_i c_i^\dagger \tau_z c_i \end{aligned} \tag{6.56}$$

where τ_z is the Paul matrix defined in terms of the two sublattices. When $\beta = 0$ and π, the Zeeman field vanishes, but the system stays in two topologically distinct phases. As β increases from 0 to π, the system varies smoothly from a trivial to a topological insulator without closing the bulk energy gap.

Setting $t_0 = 1$, $\lambda_{\rm SO} = 0.125$, and $m = 0.5$, we first explore the behavior of the B

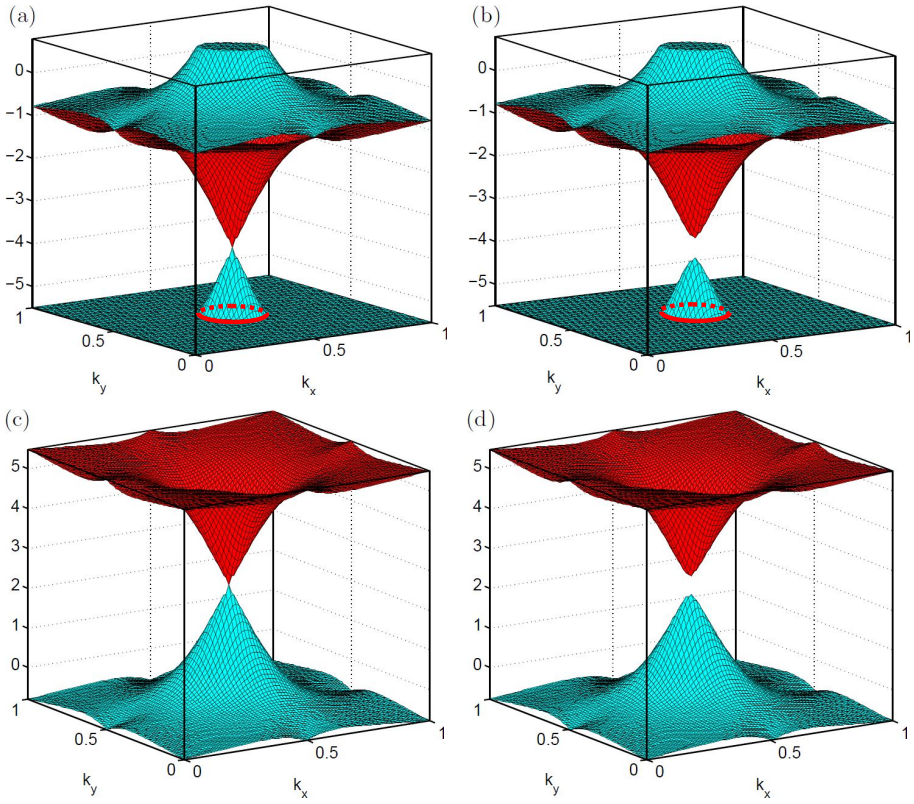

Figure 6.3: 3D plots of the two eigenvalues (marked in cyan and red) of $B(k_x, k_y) = i \ln U(k_x, k_y)$ for the Fu-Kane-Mele model at half filling: (a) when the system is a TI, i.e., $\beta = \pi$, with the branch choice taken as $(-5.5, 0.783]$; (b) when $\beta = 0.95\pi$, with the branch choice $(-5.5, 0.783]$; (c) when $\beta = \pi$, with the branch choice $(-0.785, 5.498]$; and (d) when $\beta = 0.95\pi$, with the branch choice $(-0.785, 5.498]$. The vortex loops in (a) and (b) are marked by red circles. The wavevectors k_x and k_y are defined in units of 2π.

matrix (Eq. (6.10)) in the (k_x, k_y) plane. As shown in Fig. 6.3(a), when the system is in the $\mathbb{Z}_2$-odd phase ($\beta = \pi$), for a given branch choice (from -5.5 to 0.783) there is a single vortex loop surrounding one of the TRIM (π, π). Within the loop, one of the eigenvalues of B (in cyan) is shifted by 2π, while the other eigenvalue remains continuous. Moreover, as a result of $\mathcal{T}$ symmetry, the two eigenvalues of B are degenerate at each TRIM, leading to quantized Berry phases as discussed in Sec. 6.4.3. If $\mathcal{T}$ symmetry is broken, e.g., when $\beta = 0.95\pi$, the two eigenvalues of B are no longer degenerate, even though the vortex loop is still present for such a branch choice, as shown in Fig. 6.3(b).

As discussed in Sec. 6.4.1, when $\mathcal{T}$ symmetry is preserved and the system is in the

$\mathbb{Z}_2$-odd phase, a vortex loop has to be present regardless of the branch choice. The best one can do is to compress the vortex loop to one of the TRIM in the 2D plane. This is illustrated in Fig. 6.3(c). When the branch choice is taken as $(-0.785, 5.498]$ for the $\mathbb{Z}_2$-odd case ($\beta = \pi$), the vortex loop is compressed to the point (π, π) in the (k_x, k_y) plane. On the other hand, using the same branch choice, the vortex loop can be removed if $\mathcal{T}$ symmetry is broken ($\beta = 0.95\pi$), which is shown in Fig. 6.3(d).

Using the method developed in Sec. 6.3 and Sec. 6.4 , we have calculated the total axion response (θ) along the path from $\beta = 0$ to π by summing over θ_{BK}, θ_{GD} and θ_{VL}. We would like to first explain the specific procedures for these calculations before discussing any specific results. The parallel-transport technique [52] is heavily used in these procedures, and we refer the readers to Appendix 6.7.1 for details about the parallel transport. As discussed in previous sections, the basic idea is that we first construct a smooth gauge in the bulk BZ, which is periodic only in the k_x and k_y directions. Then we can extract the unitary matrix $U(\mathbf{k}_x, \mathbf{k}_y)$ describing the gauge discontinuity (Eq. (6.8)) by calculating the overlap between the Bloch states in the top-plane and bottom-plane gauges. The logarithm of $U(\mathbf{k}_x, \mathbf{k}_y)$ with a certain branch choice would give us the B matrix. We also need to calculate the Berry curvature and Berry connections either in the top-plane gauge or in the bottom-plane gauge. Then all the formulae derived in previous sections can be applied.

To be specific, first we need to construct a smooth and periodic gauge on an arbitrary (k_x, k_y) plane, say, the $k_z = 0$ plane. In order to do this, we first construct the parallel-transport gauge (see Appendix 6.7.1) along the k_y direction at $k_x = 0$, then make a set of separate parallel transports from $k_x = 0$ to $k_x = 2\pi$ at each k_y, leaving some gauge discontinuity at the line $k_x = 2\pi$ denoted by $Y(k_y) = e^{-iD(k_y)}$. We then apply a local (in $\mathbf{k}$ space) unitary transformation $R(k_x, k_y) = e^{ik_x D(k_y)/2\pi}$ to the occupied states at each (k_x, k_y) to fix this discontinuity. In the above operation, we have insisted the smoothness of the occupied states because the R matrix is defined to be smooth

in the 2D plane ($k_z = 0$); in the meanwhile the gauge discontinuity at the boundary line ($k_x = 2\pi$) is removed. After these operations, we have successfully constructed a smooth and periodic gauge in the $k_z = 0$ plane.

Taking the smooth and periodic gauge in the $k_z = 0$ plane as a "reference gauge", at each (k_x, k_y) we further carry out two sets of parallel transports along the positive and negative k_z directions from $k_z = 0$ to $k_z = \pm\pi$. However, now the periodicity condition in k_z is relaxed so that the states are as aligned to each other as possible in the k_z direction. This makes the numeric convergence of the bulk integral (Eq. (6.6)) much easier. After such operations, we have constructed a gauge which is smooth everywhere in the bulk BZ, and in the meanwhile remain periodic in the k_x and k_y directions. Some gauge discontinuity is left at the plane $k_z = \pm\pi$, which is described by the U matrix as introduced in Sec. 6.2.2. Now we are well prepared to apply the formulae derived in Sec. 6.3 and Sec. 6.4 to specific physical systems.

The above procedures have to be implemented with caution if the system has $\mathcal{T}$ symmetry and is in the $\mathbb{Z}_2$-odd phase. In this case, it is desirable to construct a bulk gauge respecting $\mathcal{T}$ symmetry, so that both θ_{BK} and θ_{GD} vanish, and the only contribution is from θ_{VL}. For a 3D strong TI, however, the 2D $\mathbb{Z}_2$ indices for the $k_z = 0$ plane and the $k_z = \pi$ plane must be opposite. On the other hand, it is well known that it is impossible to construct a smooth and periodic gauge respecting $\mathcal{T}$ symmetry in the 2D BZ with an odd $\mathbb{Z}_2$ index [56]. Therefore, for a 3D strong TI, one has to first select the $\mathbb{Z}_2$-even plane and construct a (smooth and periodic) $\mathcal{T}$-symmetric gauge on that plane as the reference gauge. As various numeric methods for computing $\mathbb{Z}_2$ indices have been proposed, and some of them have already been implemented in first-principles code packages [130, 167], computing $\mathbb{Z}_2$ indices by itself (even in the absence of $\mathcal{P}$ symmetry) is not difficult.

The axion response θ for the FKM model is shown in Fig. 6.4. As β increase from 0 to π, the system evolves from a $\mathbb{Z}_2$-even to a $\mathbb{Z}_2$-odd phase without closing

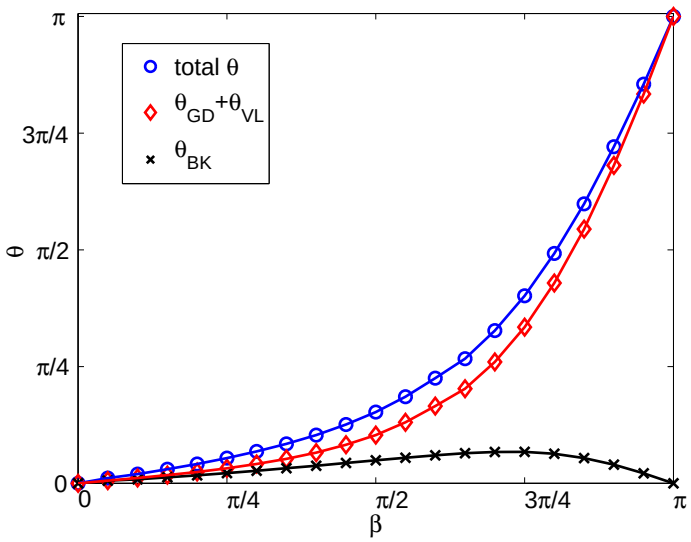

Figure 6.4: The axion response θ for the Fu-Kane-Mele model. The blue circle denotes the total response. The red-diamond curve indicates the contribution from the gauge discontinuity, including both the 2D surface integral ($\theta_{\rm GD}$) and the 1D vortex-loop integral ($\theta_{\rm VL}$). The black crosses represent the contribution from the bulk integral without enforcing periodicity in the k_z direction.

the bulk energy gap. As a result, θ increases smoothly from 0 to π. It is helpful to decompose the total θ into the bulk-BZ integral $\theta_{\rm BK}$ and the rest $\theta_{\rm GD}+\theta_{\rm VL}$, which are indicated by black crosses and red diamonds respectively. One finds that as β increases , $\theta_{\rm GD}+\theta_{\rm VL}$ becomes more and more dominating. Eventually when $\beta=\pi$, θ is completely contributed by the vortex-loop term and equals to π, because both $\theta_{\rm GD}$ and $\theta_{\rm BK}$ vanish due to a $\mathcal{T}$-symmetric gauge choice in the bulk.

It should be noted that none of the three terms $\theta_{\rm BK}$, $\theta_{\rm GD}$ and $\theta_{\rm VL}$, is independently gauge invariant. As the size of the vortex loop is dependent on the branch choice, in general both $\theta_{\rm VL}$ and $\theta_{\rm GD}$ are branch-choice dependent unless there is any additional constriant from special symmetries (e.g., $\mathcal{T}$ symmetry). However, for a given gauge choice in the bulk BZ, the bulk integral $\theta_{\rm BK}$ is fixed, therefore the sum of $\theta_{\rm VL}$ and $\theta_{\rm GD}$ should be invariant.

The above statement is demonstrated in Fig. 6.5. The blue diamonds (black plus signs) in Fig. 6.5 denote the difference between the values of $\theta_{\rm GD}$ ($\theta_{\rm VL}$) calculated

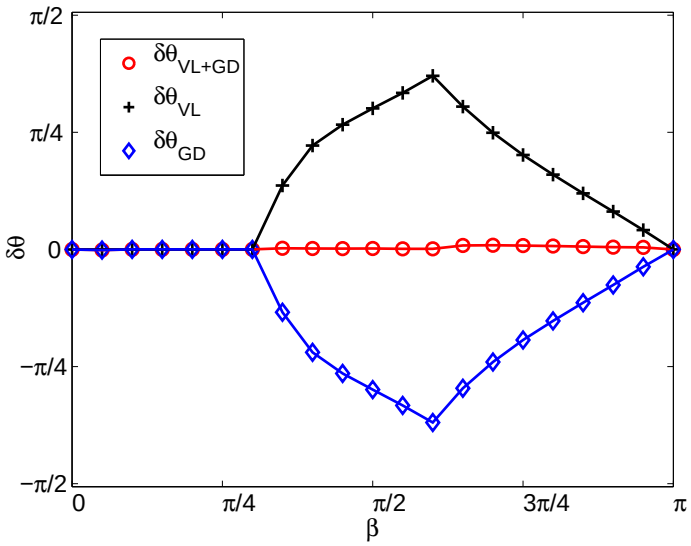

Figure 6.5: The difference between the θ values calculated with two different branch choices (see text) for the Fu-Kane-Mele model. The blue diamonds, black plus signs, and red circles denote the differences for $\theta_{\rm GD}$, $\theta_{\rm VL}$ and $\theta_{\rm VL}+\theta_{\rm GD}$.

using two different branch choices $(-2\pi, 0]$ and $(-5.5, 0.783]$. For the first branch choice ($(-2\pi, 0]$), a vortex loop starts showing up when $\beta = 0.35\pi$; for the other branch choice, however, the eigenvalues of B remain continuous throughout the 2D $\mathbf{k}$ plane until $\beta = 0.65\pi$. It is clearly shown from Fig. 6.5 that both $\theta_{\rm VL}$ and $\theta_{\rm GD}$ are branch-choice dependent. On the other hand, the red circles in Fig. 6.5 represent the difference (between the two branch choices) for $\theta_{\rm GD}+\theta_{\rm VL}$, which remains vanishingly small throughout the adiabatic path. It thus numerically tested that the sum of $\theta_{\rm VL}$ and $\theta_{\rm GD}$ is invariant against different branch choices.

For a given branch choice, however, there is still freedom in choosing the bulk gauge. In this regard, both $\theta_{\rm BK}$ and $\theta_{\rm GD}+\theta_{\rm VL}$ are dependent on the choice of the bulk gauge. However, since the gauge choice in the bulk is constructed in such a way that in the k_z direction they remain as aligned to each other as possible [4], the bulk integral $\theta_{\rm BK}$ is typically small. This explains why $\theta_{\rm GD}+\theta_{\rm VL}$ dominates over $\theta_{\rm BK}$ in Fig. 6.4.

[4]We define two set of states at neighboring $\mathbf{k}$ points $|u_{n\mathbf{k}}\rangle$ and $|u_{n\mathbf{k}+\delta\mathbf{k}}\rangle$ to be optimally aligned if the overlap matrix $M_{mn}(\mathbf{k}) = \langle u_{m\mathbf{k}}|u_{n\mathbf{k}+\delta\mathbf{k}}\rangle$ is Hermitian, which can be obtained by a parallel transport operation. If the states are optimally aligned everywhere in BZ, which can be obtained by parallel transport without enforcing periodicity, the Berry connection vanishes everywhere, leading to a vanishing bulk integral $\theta_{\rm BK}$.

6.6 Summary

To summarize, we have developed a new method for computing the θ term. The basic idea is to relax the periodicity condition for the gauge in one $\mathbf{k}$ direction, thus introducing some gauge discontinuity resided at a 2D planar surface. The total contribution is then the sum of the 3D integral over the bulk BZ excluding the gauge-discontinuity plane (θ_{BK}) and the contribution from the gauge discontinuity θ_{GD}, which is expressed as a 2D integral over the 2D $\mathbf{k}$ plane as shown in Eq. (6.16) and Eq. (6.31). Furthermore, depending on the branch choice of $B(\mathbf{k})$ and the topological property of the system, the 2D $\mathbf{k}$ plane may be further divided into subregions by 1D vortex loops. In this case one must also consider the vortex-loop contribution as expressed in Eq. (6.52). The total θ is then $\theta = \theta_{\mathrm{BK}} + \theta_{\mathrm{GD}} + \theta_{\mathrm{VL}}$.

As the periodicity condition in one $\mathbf{k}$ direction (k_z) is relaxed, the states do not twist as strong as in the case when both periodicity and smoothness are required. This leads to much easier numeric convergence of the 3D bulk integral (Eq. (6.6)). The lost of periodicity is compensated by extra contributions from gauge discontinuity (θ_{GD}) and possible vortex loops (θ_{VL}). The formulae for both terms turn out to be simple, which can be numerically implemented without much difficulty.

It is interesting to note that if a gauge respecting $\mathcal{T}$ symmetry has been constructed in the bulk BZ for a $\mathcal{T}$-invariant system, then both θ_{BK} and θ_{GD} vanish. The only surviving term θ_{VL} is proved to be either 0 or π, depending on the topology of the system. Our theory thus provides a new interpretation to the (formally) quantized magnetoelectric response in TIs.

We have applied our method to the Fu-Kane-Mele model with applied staggered Zeeman field. We calculated the axion response for the model along a path with broken $\mathcal{T}$ symmetry connecting the $\mathbb{Z}_2$-even and $\mathbb{Z}_2$-odd phases. Our results agree well with the previous results obtained from other methods [20, 168]. In particular, we find that

the gauge-discontinuity contribution ($\theta_{\mathrm{GD}} + \theta_{\mathrm{VL}}$) becomes more and more dominating as the system approaches the $\mathbb{Z}_2$-odd phase. In the TI phase, as mentioned above,θ is completely contributed by the vortex-loop term for a $\mathcal{T}$-symmetric gauge in the bulk BZ, and the π quantization of θ is due to the π quantization of the Berry phases around a single vortex loop.

Our method can also be generalized in such a way that a 3D BZ is divided into multiple subvolumes. The subvolumes may meet each other at multiple 2D boundaries, and these boundaries meet at curves which may be vortex lines. In such a more complicated case, the formula for θ_{GD} still applies, but the definition of a vortex loop as given in Sec. 6.4.1 has to be generalized to fit a situation that the two adjacent B matrices around a vortex loop no longer commute with each other. Thus the formula for θ_{VL} may need to be modified. We leave this for future study.

From theoretical point of view, the results presented in this chapter is a step forward in understanding the orbital Chern-Simons ME coupling. We introduced the gauge-discontinuity and vortex-loop contributions to θ, and derived formulae for them. We illustrated that the π quantization of θ in TIs may be explained by the quantization of Berry phases around vortex loops enclosing degeneracy points at the TRIM.

From the perspective of first-principles calculations, our method is numerically efficient. It can be implemented in standard first-principles code packages without much difficulty. This makes it possible to accurately compute the Chern-Simons magnetoelectric coupling for realistic materials, thus motivating the search for functional materials with giant orbital magnetoelectric coupling.

This work is supported by NSF Grants DMR-14-08838.

6.7 Appendix

6.7.1 Parallel transport

We discuss about how to carry out the parallel transport operation and construct a parallel-transport gauge starting from a set of occupied eigenstates with random gauge choice on a give $\mathbf{k}$ path. The basic idea is to make the (periodic part of) Bloch states at a particular point on a $\mathbf{k}$ path to be as aligned as possible to the states of the previous $\mathbf{k}$ point. If the $\mathbf{k}$ path is chosen to be closed, the Bloch states after the parallel-transport operation may be different from the states in the original gauge by a phase factor, which is indeed the Berry phase. The Berry phase accumulated along the path can be gradually smeared out by distributing the phase to every $\mathbf{k}$ point on the path. After this operation, one obtains a gauge which is both smooth and periodic along the path, and is called a parallel-transport gauge.

To be more specific, let us consider a set of occupied bands $|u_{n\mathbf{k}}\rangle$, with $n = 1, ...N$, which are isolated from other bands in energy everywhere in the BZ . Let us take a closed $\mathbf{k}$ path along k_z running from 0 to 2π, and the path is sampled by J discrete points , so that $\mathbf{k}_j = 2\pi(j-1)\hat{z}/J$. Assume that the eigenstates with some random gauge $|u^0_{n\mathbf{k}_j}\rangle$ have been obtained for $j = 1, ...J$, and a periodic gauge is chosen at the $J+1$th point so that $|u^0_{n\mathbf{k}_{J+1}}\rangle = e^{-i2\pi z}|u^0_{n\mathbf{k}_1}\rangle$. To carry out the parallel transport, we need to make the overlap matrix between the occupied states at $\mathbf{k}_{j+1}$ and $\mathbf{k}_j$, i.e., $M_{mn}(j) = \langle u^0_{m\mathbf{k}_j}|u^0_{n\mathbf{k}_{j+1}}\rangle$, Hermitian. This can be done as follows. At each $\mathbf{k}_j$, make a singular-value decomposition to the overlap matrix $M_j = V_j\Sigma_j W_j^\dagger$, where V and W are unitary and Σ is Hermitian. Then apply a unitary transformation $R_j = W_j V_j^\dagger$ to $|u^0_{n\mathbf{k}_{j+1}}\rangle$: $|\widetilde{u}_{n\mathbf{k}_{j+1}}\rangle = \sum_{m=1}^N R_{j,mn}|u^0_{m\mathbf{k}_{j+1}}\rangle$. Now the overlap matrix between the rotated states at neighboring $\mathbf{k}$ points becomes Hermitian. If one keeps doing such an operation from $j = 1$ to $j = J$, the states would remain as aligned as possible to each other at all $\mathbf{k}$ points on the path. However, there is some gauge discontinuity left at the boundary:

$|\widetilde{u}_{n\mathbf{k}_{J+1}}\rangle = e^{-i2\pi z}\sum_m \Lambda_{mn}|\widetilde{u}_{m\mathbf{k}_1}\rangle$, where Λ is a unitary matrix. The logarithm of the eigenvalues of Λ: $\beta_n = -i\ln\lambda_n$ can be considered as the non-Abelian Berry phase of the states.

The above gauge after parallel transport is smooth along the $\mathbf{k}$ path, but not periodic. To restore periodicity, we need to rotate all the states on the $\mathbf{k}$ path to the basis that diagonalizes Λ: $|u'_{n\mathbf{k}_j}\rangle = \sum_{m=1}^{N}|\widetilde{u}_{m\mathbf{k}_j}\rangle\, L_{mn}$, where L is the eigenvector matrix of Λ. Then we gradually smear out the discontinuity by applying the following phase twist to the states at $\mathbf{k}_j$: $|u_{n\mathbf{k}_j}\rangle = e^{-i(j-1)\beta_n/J}|u'_{n\mathbf{k}_j}\rangle$. Then we have constructed a gauge that is both smooth and periodic along the $\mathbf{k}$ path, and is called a parallel-transport gauge.

6.7.2 Integration over λ in the formula for $\theta_{\mathbf{GD}}$

Suppose that $A_x^{(0)}(\mathbf{k})$, $A_y^{(0)}(\mathbf{k})$, $\Omega_{xy}^{(0)}(\mathbf{k})$, $B(\mathbf{k})$, $B_x(\mathbf{k})$, and $B_y(\mathbf{k})$ are known. Then our job is to compute the quantities in Eqs. (6.32-6.34) and plug them into Eq. (6.31). These three quantities can all be computed analytically in the sense that we do not have to discretize the λ axis. The plan is as follows. The first term in Eq. (6.31) is independent of λ and is trivial. For the remaining terms, at each $\mathbf{k}$, locally diagonalize $B(\mathbf{k})$, transform all of the matrices $A_x^{(0)}(\mathbf{k})$, $A_y^{(0)}(\mathbf{k})$, $B_x(\mathbf{k})$, and $B_y(\mathbf{k})$ to the basis that locally diagonalizes $B(\mathbf{k})$, i.e.,

$$\begin{aligned}
&B(\mathbf{k}) \to V^\dagger(\mathbf{k})\, B(\mathbf{k})\, V(\mathbf{k}),\\
&B_{x(y)}(\mathbf{k}) \to V^\dagger(\mathbf{k})\, B_{x(y)}(\mathbf{k})\, V(\mathbf{k}),\\
&A_{x(y)}(\mathbf{k}) \to V^\dagger(\mathbf{k})\, A_{x(y)}(\mathbf{k})\, V(\mathbf{k}),\\
&\Omega_{xy}(\mathbf{k}) \to V^\dagger(\mathbf{k})\, \Omega_{xy}(\mathbf{k})\, V(\mathbf{k})\ ,
\end{aligned} \tag{6.57}$$

where $V(\mathbf{k})$ is the eigenvector matrix of $B(\mathbf{k})$. Then one can compute the trace in this basis. Letting $B_{mn} = b_n\,\delta_{mn}$, we find

$$\begin{aligned}
\overline{B}_{x,mn}(\lambda) &= \int_0^\lambda d\mu\; e^{-i\mu b_m}\, B_{x,mn}\, e^{i\mu b_n}\\
&= g_{mn}(\lambda)\, B_{x,mn}\ ,
\end{aligned} \tag{6.58}$$

where

$$g_{mn}(\lambda) = \frac{e^{-i\lambda(b_m - b_n)} - 1}{-i(b_m - b_n)} \; . \tag{6.59}$$

Then

$$\begin{aligned} \overline{\overline{B}}_{x,mn} &= \Big(\int_0^1 g_{mn}(\lambda)\, d\lambda \Big)\, B_{x,mn} \\ &= \Big(\frac{e^{i(b_n - b_m)} - 1}{-(b_n - b_m)^2} - \frac{1}{i(b_n - b_m)} \Big)\, B_{x,mn} \; , \end{aligned} \tag{6.60}$$

and

$$\overline{\overline{B}}_{x,y,mn} = i \sum_l \Big(\int_0^1 g_{ml}(\lambda)\, g_{ln}(\lambda)\, d\lambda \Big) \Big(B_{x,ml}\, B_{y,ln} - B_{y,ml}\, B_{x,ln} \Big) \; . \tag{6.61}$$

Because we are interested in the trace of $B\,\overline{\overline{B}}_{x,y}$ in the basis that B is locally diagonal, only the diagonal matrix elements of $\overline{\overline{B}}_{x,y}$ are relevant. After carrying out the integral in Eq. (6.61) one obtains the following expression:

$$\overline{\overline{B}}_{x,y,nn} = i \sum_m \Big(\frac{2}{(b_n - b_m)^2} - \frac{2\sin(b_m - b_n)}{(b_m - b_n)^3} \Big) \Big(B_{x,nm} B_{y,mn} - B_{y,nm} B_{x,mn} \Big) \; . \tag{6.62}$$

If two eigenvalues b_m and b_n are degenerate, one needs to take the limit $(b_n - b_m) \to 0$. It turns out that both quantities are finite:

$$\lim_{b_n \to b_m} \overline{\overline{B}}_{x,mn} = B_{x,mn}/2 \; , \tag{6.63}$$

and

$$\lim_{b_m \to b_n} \overline{\overline{B}}_{x,y,nn} = \frac{i}{3} \Big(B_{x,nm} B_{y,mn} - B_{y,nm} B_{x,mn} \Big) \; . \tag{6.64}$$

Of course the entire calculation still has to be done on a discretized mesh on the $\mathbf{k}$ plane, with finite-difference expressions used to evaluate objects like $A_x^{(0)}(\mathbf{k})$, so it it not "exact." But it is at least nice that we eliminate the need to discretize the λ axis, instead doing all λ integrals analytically.

6.7.3 Derivation of Eq. (6.41)

Let us prove Eq. (6.41). From Eq. (6.25), Eq. (6.26) and Eq. (6.23), we know that:

$$A^{(\lambda)}_{x(y)} = \widetilde{A}^{(\lambda)}_{x(y)} + \Gamma_{x(y)}(0,\lambda) \tag{6.65}$$

where

$$\widetilde{A}^{(\lambda)}_{x(y)} = W^{\dagger}(\lambda)\, A^{(0)}_{x(y)}\, W(\lambda)\ , \tag{6.66}$$

and the function $\Gamma_{x(y)}(\lambda_1,\lambda_2)$ is defined as

$$\Gamma_{x(y)}(\lambda_1,\lambda_2) = \int_{\lambda_1}^{\lambda_2} d\mu\, W^{\dagger}(\mu)\, B_{x(y)}\, W(\mu)\ . \tag{6.67}$$

Let $\lambda = 1$, we get the expression of $A^{(1)}_{x(y)}$

$$A^{(1)}_{x(y)} = \widetilde{A}^{(1)}_{x(y)} + \Gamma_{x(y)}(0,1) \tag{6.68}$$

Apply a unitary transformation $W(1-\lambda)$ to the matrix $A^{(1)}_{x(y)}$, one obtains the following:

$$\begin{aligned} W(1-\lambda)\, A^{(1)}_{x(y)}\, W^{\dagger}(1-\lambda) &= \widetilde{A}^{(\lambda)}_{x(y)} + \Gamma_{x(y)}(\lambda-1,\lambda)\ , \\ &= A^{(\lambda)}_{x(y)} + \Gamma_{x(y)}(\lambda-1,0)\ , \end{aligned} \tag{6.69}$$

where a variable transformation $(\lambda+\nu-1)\to\mu$ has been made to obtain the second term ($\Gamma_{x(y)}(\lambda-1,\lambda)$) on the RHS of the first line in Eq. (6.69). The integral from $\lambda-1$ to λ in $\Gamma_{x(y)}(\lambda-1,\lambda)$ is further divided into two integrals: one from $\lambda-1$ to 0, and the other from 0 to λ. $A^{(\lambda)}_{x(y)}$ in the second line is then obtained by combining the integral from 0 to λ together with $W^{\dagger}(\lambda)\, A^{(0)}_{x(y)}\, W(\lambda)$ (Eq. (6.65)). Therefore:

$$A^{(\lambda)}_{x(y)} = W(1-\lambda)\, A^{(1)}_{x(y)}\, W^{\dagger}(1-\lambda)\ - \Gamma_{x(y)}(\lambda-1,0)\ . \tag{6.70}$$

It immediately follows that:

$$\begin{aligned} A^{(1-\lambda)}_{x(y)} &= W(\lambda)\, A^{(1)}_{x(y)}\, W^{\dagger}(\lambda) - \Gamma_{x(y)}(-\lambda,0) \\ &= W(\lambda)\, A^{(1)}_{x(y)}\, W^{\dagger}(\lambda) - \int_0^{\lambda} d\mu\, W(\mu)\, B_{x(y)}\, W^{\dagger}(\mu)\ , \end{aligned} \tag{6.71}$$

where we let $\mu \to -\mu$ in going from the first to the second line in Eq. (6.71). Combining Eq. (6.38)-(6.39) and Eq. (6.71), Eq. (6.41) can be proved in a straightforward manner:

$$\begin{aligned} A^{(1-\lambda)}_{x(y)}(-\mathbf{k}) &= e^{-i\lambda B(-\mathbf{k})}\, A^{(1)}_{x(y)}(-\mathbf{k})\, e^{i\lambda B(-\mathbf{k})} - \int_0^\lambda d\mu\, e^{-i\mu B(-\mathbf{k})}\, B_{x(y)}(-\mathbf{k})\, e^{i\mu B(-\mathbf{k})} \\ &= \sigma_y\, e^{-i\lambda B^T(\mathbf{k})} \left(A^{(0)}_{x(y)}(\mathbf{k})\right)^T e^{i\lambda B^T(\mathbf{k})}\, \sigma_y + \sigma_y \int_0^\lambda d\mu\, e^{-i\mu B^T(\mathbf{k})}\, B^T_{x(y)}(\mathbf{k})\, e^{i\mu B^T(\mathbf{k})}\, \sigma_y \\ &= \sigma_y \left(W^\dagger(\lambda)\, A_{x(y)}(\mathbf{k})^{(0)}\, W(\lambda) + \int_0^\lambda d\mu\, W^\dagger(\mu)\, B_{x(y)(\mathbf{k})\, W(\mu)}\right)^T \sigma_y\ . \end{aligned} \tag{6.72}$$

The last line in Eq. (6.72) is simply $\sigma_y \left(A^{(\lambda)}_{x(y)}(\mathbf{k})\right)^T \sigma_y$, thus Eq. (6.41) is proved. It is confirmed that $\theta_{\rm GD}$ vanishes for a $\mathcal{T}$-invariant gauge.

6.7.4 Derivation of Eq. (6.48)

To derive Eq. (6.48), it is convenient to decompose $G(\mathbf{k})$ into four terms G_1, G_2, G_3 and G_4, corresponding to the four terms on the right-hand side (RHS) of Eq. (6.31):

$$G_1 = B\, \Omega^{(0)}_{xy} \tag{6.73}$$

$$G_2 = iB\, [\, \overline{B}_x(\lambda), \overline{B}_y(\lambda)\,] \tag{6.74}$$

$$G_3 = iB\, [\, A^{(0)}_x, \overline{B}_y(\lambda)\,] \tag{6.75}$$

$$G_4 = iB\, [\, \overline{B}_x(\lambda), A^{(0)}_y\,] \tag{6.76}$$

Since all the quantities e.g. Ω_{xy}, $A_{x(y)}$ are defined in the bottom-plane gauge, we will drop the superscript "(0)" (indicating the bottom-plane gauge) in later steps. We also define the primed quantities, e.g., the Berry connections and the Berry curvature as

$$A'_{x(y)} = V^\dagger\, A_{x(y)}\, V\ , \tag{6.77}$$

$$\Omega'_{xy} = V^\dagger\, O_{xy}\, V\ , \tag{6.78}$$

$$\overline{B}'_{x(y)} = V^\dagger\, \overline{B}_{x(y)}\, V\ . \tag{6.79}$$

We will prove Eq. (6.48) by explicitly calculating the four terms in Eq. (6.73)-(6.76).

The G_1 term

Plug Eq. (6.47) first into the expression of G_1 (Eq. (6.73)), one obtains the following:

$$\begin{aligned} \mathrm{Tr}\Big[\, G_1 \,\Big] &= \mathrm{Tr}\Big[\, V\,\Delta_1\,V^\dagger \Omega_{xy}\, V\, V^\dagger \,\Big] \\ &= \mathrm{Tr}\Big[\, \Delta_1\, V^\dagger \Omega_{xy}\, V \,\Big] \end{aligned} \tag{6.80}$$

Note that $\Omega'_{xy} = V^\dagger\,\Omega_{xy}\,V$ is associated with the Berry curvature of the Bloch states at the bottom plane that are unitarily transformed by V: $|\overline{u}^{(0)}\rangle = |u^{(0)}\rangle\, V$. One can express the Berry curvature of $|\overline{u}^{(0)}\rangle$ (denoted by $\overline{\Omega}_{xy}$) in terms of $\mathbf{A}_x$, $\mathbf{A}_y$, Ω_{xy}, V and the partial derivatives of V,

$$\overline{\Omega}_{xy} = \Omega'_{xy} + \Lambda_{xy} + i[\, C_x, A'_y\,] - i[\, C_y, A'_x\,] \tag{6.81}$$

where

$$C_{x(y)} = iV^\dagger\,\partial_{x(y)} V\ , \tag{6.82}$$

$$\Lambda_{xy} = \partial_x C_y - \partial_y C_x\ , \tag{6.83}$$

can be considered as the Berry connection and Berry curvature in the "gauge space".

From Eq. (6.81) it immediately follows

$$\mathrm{Tr}\Big[\, G_1 \,\Big] = \mathrm{Tr}\Big[\, \Delta_1\,\overline{\Omega}_{xy} - \Delta_1\,\Lambda_{xy} - i\Delta_1\,[\, C_x, A'_y\,] + i\Delta_1\,[\, C_y, A'_x\,]\,\Big]\ . \tag{6.84}$$

Before further simplifying Eq. (6.84), let us go to the other terms and come back to G_1 later.

The G_3 and G_4 terms

Let us deal with the G_3 and G_4 terms. Since B_x and B_y are involved in G_3 and G_4, let us first evaluate these two terms.

$$\begin{aligned} B_x &= \partial_x(\, V\,\Delta_1\,V^\dagger\,) \\ &= \partial_x V\,\Delta_1\,V^\dagger + V\,\Delta_1\,\partial_x V^\dagger \\ &= iV\,[\,\Delta_1, C_x\,]\,V^\dagger \end{aligned} \tag{6.85}$$

Similarly, $B_y = iV\,[\,\Delta_1, C_y\,]\,V^\dagger$. Plug the expression of B_x and B_y into Eq. (6.23), one immediately obtains the following equation:

$$\overline{B}_{x(y)}(\lambda) = \int_0^\lambda du\, V\, e^{-iu\Delta_1}\, i[\,\Delta_1, C_{x(y)}\,]\, e^{iu\Delta_1}\, V^\dagger\,. \tag{6.86}$$

In the basis that locally diagonalizes $B_0,\ \Delta B \to V^\dagger\,\Delta B\, V = \Delta_1,\ A_{x(y)} \to A'_{x(y)},$ $\overline{B}_{x(y)} \to \overline{B}'_{x(y)}$, we get the following expression for $\mathrm{Tr}\Big[\,G_3\,\Big]$:

$$\begin{aligned}
\mathrm{Tr}\Big[\,G_3\,\Big] &= \mathrm{Tr}\Big[\,i\Delta_1\,[\,A'_x,\,\overline{B}'_y,]\Big] \\
&= \mathrm{Tr}\Big[\int_0^\lambda du\, i\Delta_1\,[\,A'_x,\, e^{-iu\Delta_1}\, i[\,\Delta_1, C_y\,]\, e^{iu\Delta_1}\,]\Big] \\
&= \mathrm{Tr}\Big[\int_0^\lambda du\, iA'_x\,[\,e^{-iu\Delta_1}\, i[\,\Delta_1, C_y\,]\, e^{iu\Delta_1},\, \Delta_1\,]\Big] \\
&= \mathrm{Tr}\Big[\,iA'_x \int_0^\lambda du\, \partial_u\Big(\,e^{-iu\Delta_1}\,[\,\Delta_1, C_y\,]\, e^{iu\Delta_1}\Big)\Big] \\
&= \mathrm{Tr}\Big[\,iA'_x\, e^{-i\lambda\Delta_1}\,[\,\Delta_1, C_y\,]\, e^{i\lambda\Delta_1} - iA'_x\,[\,\Delta_1, C_y\,]\Big]\,,
\end{aligned} \tag{6.87}$$

where we have used the equation:

$$[\,e^{-iu\Delta_1}\, i[\,\Delta_1, C_y\,]\, e^{iu\Delta_1},\, \Delta_1\,] = \partial_u\Big(\,e^{-iu\Delta_1}\,[\,\Delta_1, C_y\,]\, e^{iu\Delta_1}\Big) \tag{6.88}$$

in going from the third to the fourth line in Eq. (6.87). Taking use of the cyclic property of trace, one immediately realizes that the second term in the last line of Eq. (6.87) cancels the last term on the RHS of Eq. (6.84), which will be dropped in later steps. Therefore,

$$\begin{aligned}
\int_0^1 d\lambda\, \mathrm{Tr}\Big[\,G_3\,\Big] &= \int_0^1 d\lambda\, \mathrm{Tr}\Big[\,iA'_x\, e^{-i\lambda\Delta_1}\,[\,\Delta_1, C_y\,]\, e^{i\lambda\Delta_1}\,\Big] \\
&= \int_0^1 d\lambda\, \mathrm{Tr}\Big[\,-A'_x\,\partial_\lambda(\,e^{-i\lambda\Delta_1}\, C_y\, e^{i\lambda\Delta_1}\,)\Big] \\
&= \mathrm{Tr}\Big[\,-A'_x\,(\,e^{-i\lambda\Delta_1}\, C_y\, e^{i\lambda\Delta_1}\,)|_{\lambda=0}^{\lambda=1}\Big] \\
&= 0\,,
\end{aligned} \tag{6.89}$$

where the following equation is used to go from the second to the third line in Eq. (6.89):

$$ie^{-i\lambda\Delta_1}\,[\,\Delta_1, C_y\,]\, e^{i\lambda\Delta_1} = -\partial_\lambda(\,e^{-i\lambda\Delta_1}\, C_y\, e^{i\lambda\Delta_1}\,)\,. \tag{6.90}$$

Similar derivations can be applied to the G_4 term:

$$\mathrm{Tr}\Big[\, G_4 \,\Big] = \mathrm{Tr}\Big[\, iA'_y\, e^{-i\lambda\Delta_1}\,[\, C_x, \Delta_1\,]\, e^{i\lambda\Delta_1} - iA'_y\,[\, C_x, \Delta_1\,]\,\Big]\,. \tag{6.91}$$

The second term on the RHS of Eq. (6.91) cancels the third term on the RHS of Eq. (6.84). Drop the second term in Eq. (6.91) and integrate over λ, one obtains $\int_0^1 d\lambda\,\mathrm{Tr}\Big[\, G4\,\Big] = 0$

The G_2 term

In the basis that locally diagonalizes B_0 ,

$$\mathrm{Tr}\Big[\, G_2 \,\Big] = \mathrm{Tr}\Big[\, i\Delta_1\,[\, \overline{B}'_x, \overline{B}'_y\,]\,\Big]\,. \tag{6.92}$$

On the other hand, combining Eq. (6.86), Eq. (6.79) and Eq. (6.90), we get the following expression for $\overline{B}'_{x(y)}$:

$$\begin{aligned} \overline{B}'_{x(y)} &= -\int_0^\lambda d\mu\partial_u(\, e^{-iu\Delta_1}\, C_{x(y)}\, e^{iu\Delta_1}\,) \\ &= C_{x(y)} - \widetilde{C}_{x(y)}\ , \end{aligned} \tag{6.93}$$

where $\widetilde{C}_{x(y)} = e^{-i\lambda\Delta_1}\, C_{x(y)}\, e^{i\lambda\Delta_1}$. It follows that:

$$\mathrm{Tr}\Big[\, G_2 \,\Big] = \mathrm{Tr}\Big[\, i\Delta_1\,[\, \widetilde{C}_x - C_x\,,\, \widetilde{C}_y - C_y\,]\,\Big]\,. \tag{6.94}$$

If one expands the RHS of Eq. (6.94), one would obtain four commutators between $C_{x(y)}$ and $\widetilde{C}_{x(y)}$. Since $e^{\pm i\lambda\Delta_1}$ commute with Δ_1, the term involving $[\,\widetilde{C}_x, \mathcal{C}_y\,]$ equals to the term with $[\, C_x, C_y\,]$. Therefore,

$$\mathrm{Tr}\Big[\, G_2 \,\Big] = \mathrm{Tr}\Big[\, \Delta_1\,\Big(\, 2i[\, C_x, C_y\,] - i[\, C_x, \widetilde{C}_y\,] - i[\,\widetilde{C}_x, C_y\,]\,\Big)\,\Big]\,. \tag{6.95}$$

The second term on the RHS of Eq. (6.95) can be written as a total derivative of λ

$$\mathrm{Tr}\Big[\, -i\Delta_1\,[\, C_x, \widetilde{C}_y\,]\,\Big] = -\mathrm{Tr}\Big[\partial_\lambda(\, e^{-i\lambda\Delta_1}\, C_y\, e^{i\lambda\Delta_1}\, C_x\,)\,\Big]\,. \tag{6.96}$$

We need to use $\Delta_1\, e^{\pm i\lambda\Delta_1} = \mp i\partial_\lambda(e^{\pm i\lambda\Delta_1})$ to obtain the above equation. Integrate Eq. (6.96) over λ, one obtains zero. Similarly, after integrating over λ, the third term on the RHS of Eq. (6.95) also vanishes. Therefore,

$$\int_0^1 d\lambda\, \mathrm{Tr}\Big[\, G_2 \,\Big] = \mathrm{Tr}\Big[\, 2\Delta_1\, i[\, C_x, C_y\,]\,\Big] \tag{6.97}$$

Note that the gauge-covariant Berry curvature defined in the gauge space $\widetilde{\Lambda}_{xy} = \Lambda_{xy} - i[\, C_x, C_y\,]$ has to vanish (Λ_{xy} defined in Eq. (6.83)). Because $\widetilde{\Lambda}_{xy}$ is the Berry curvature projected onto the unoccupied subspace, which is zero. Therefore, $\Lambda_{xy} = i[\, C_x, C_y\,]$. It can also be shown by explicitly writing out the commutator of C_x and C_y:

$$\begin{aligned} i[\, C_x, C_y\,] &= i(-V^\dagger\, \partial_x V\, V^\dagger \partial_y V + V^\dagger\, \partial_y V\, V^\dagger \partial_x V) \\ &= i(V^\dagger\, V\, \partial_x V^\dagger \partial_y V - V^\dagger\, V\, \partial_y V^\dagger \partial_x V) \\ &= i\partial_x(V^\dagger\, \partial_y V) - i\partial_y(V^\dagger\, \partial_x V) \\ &= \Lambda_{xy} \end{aligned} \tag{6.98}$$

We have used the fact that $V\, V^\dagger = 1$ and $\partial_{x(y)}(V\, V^\dagger) = 0$ in the above derivations. Therefore,

$$\int_0^1 d\lambda\, \mathrm{Tr}\Big[\, G_2 \,\Big] = \mathrm{Tr}\Big[\, 2\Delta_1\, \Lambda_{xy}\,\Big] \tag{6.99}$$

Combine Eq. (6.84), Eq. (6.87), Eq. (6.91) and Eq. (6.99), we get:

$$\begin{aligned} \theta_{\mathrm{shift}} &= \frac{-1}{4\pi}\int dk_x dk_y \int_0^1 d\lambda\, \mathrm{Tr}\Big[\, G_1 + G_2 + G_3 + G_4\,\Big] \\ &= \frac{-1}{4\pi}\int_{\mathcal{S}} dk_x dk_y\, \mathrm{Tr}\Big[\, \Delta_1\Omega'_{xy} + \Delta_1\Lambda_{xy}\,\Big] \\ &= \frac{-1}{4\pi}\int_{\mathcal{S}} dk_x dk_y \Big(\, 2\pi\,(\Omega'_{xy})_{11} + 2\pi\,(\Lambda_{xy})_{11}\,\Big) \\ &= -\Big(\,\phi_1(\mathcal{C}) + \xi_1(\mathcal{C})\,\Big)/2 \end{aligned} \tag{6.100}$$

Eq. (6.48) has been proved. Again, θ_{VL} is equal and opposite to θ_{shift}

Chapter 7

Summary and outlook

We have studied some of the properties of topological materials in this thesis . To be specific, we first carried out first-principles calculations on In- and Sb-doped Bi_2Se_3, where the effects of disorder and different orbital character have been studied. We found that the phase-transition mechanisms in the two solid solutions are distinct. In the Sb-doped case, the phase transition is dominated by a gradual decrease of the effective spin-orbit coupling strength, which is compatible with the classical "linear-gap-closure" picture. On the other hand, for the In-doped Bi_2Se_3, we found that the topological phase is suppressed at low impurity compositions by the In $5s$ orbitals and the tendency of In segregation. Moreover, the phase transition may be better described by a local percolation scenario.

We also studied the Weyl semimetals emerging from noncentrosymmetric topological insulators. We have clarified the general theory for the topological phase transitions in 3D topological insulators with broken inversion symmetry, and proved that an intermediate Weyl semimetal must show up connecting the $\mathbb{Z}_2$-odd and $\mathbb{Z}_2$-even phases. We also proposed a few materials as candidates of Weyl semimetals as discussed in Chapter 4.

As spin-orbit coupling is indispensable to obtain nontrivial band topology in most cases, we propose a quantitative definition for the band inversions driven by spin-orbit coupling in insulators, which is known as the "spin-orbit spillage". In Chapter 5, the spin-orbit spillage has been calculated in various topological systems, and turned out to be closely related to the nontrivial topological indices.

In the last chapter, we proposed a new method to calculate the Chern-Simons orbital magnetoelectric response (θ term). Based on the conventional formula for the θ term, we derived formulae for the contributions from gauge discontinuities and "vortex loops", so that the total θ consists of three terms expressed as integrals over 3D, 2D and 1D manifolds. Our method turned out to be numerically efficient, and may provide a new interpretation to the formally quantized θ in topological insulators.

Although the framework of topological band theory has been well established, the generalizations to disordered systems [169], to strongly correlated systems [49, 170, 171, 172], and to finite temperatures [173, 174] are still in the early stages. There are still a lot of unexplained experimental phenomena and open questions in this field. For example, recently it was observed by angle resolved photoelectron spectroscopy that the topological phase transitions in $(Bi_{1-x}In_x)_2Se_3$ happened with a sudden closure of bulk energy gap [175], which challenged the conventional wisdom from topological band theory. It suggests that more theoretical works are needed to shed light on the phase-transition behavior in topological systems with strong disorder. There is also a lot of interest in searching for possible topological phases in strongly correlated systems. For example, it was proposed that some kind of effective spin-orbit coupling could be generated by electron-electron interactions [50], which might lead to "topological Mott insulators" [49]. Such interaction-driven topological phase was predicted to exist in $LaNiO_3/LaAlO_3$ heterostructures [176, 143], but it has not been realized by experiments yet. Therefore, it would be both challenging and interesting to search for such interaction-generated, or interaction-enhanced topological systems. It requires further development on the theory for topological phases in strongly correlated systems, as well as a deeper understanding on the interplay between spin-orbit coupling and electron-electron interactions [177].

References

[1] M. P. Marder, Condensed matter physics. John Wiley & Sons, 2010.

[2] R. M. Martin, Electronic structure: basic theory and practical methods. Cambridge university press, 2004.

[3] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," Phys. Rev., vol. 136, pp. B864–B871, Nov 1964.

[4] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," Phys. Rev., vol. 140, pp. A1133–A1138, Nov 1965.

[5] D. H. Lee and J. D. Joannopoulos, "Simple scheme for surface-band calculations. i," Phys. Rev. B, vol. 23, pp. 4988–4996, May 1981.

[6] D. H. Lee and J. D. Joannopoulos, "Simple scheme for surface-band calculations. ii. the green's function," Phys. Rev. B, vol. 23, pp. 4997–5004, May 1981.

[7] M. L. Sancho, J. L. Sancho, J. L. Sancho, and J. Rubio, "Highly convergent schemes for the calculation of bulk and surface green functions," Journal of Physics F: Metal Physics, vol. 15, no. 4, p. 851, 1985.

[8] V. I. Anisimov, F. Aryasetiawan, and A. Lichtenstein, "First-principles calculations of the electronic structure and spectra of strongly correlated systems: the lda+ u method," Journal of Physics: Condensed Matter, vol. 9, no. 4, p. 767, 1997.

[9] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, and D. Vanderbilt, "Maximally localized wannier functions: Theory and applications," Rev. Mod. Phys., vol. 84, pp. 1419–1475, Oct 2012.

[10] J. R. Yates, X. Wang, D. Vanderbilt, and I. Souza, "Spectral and fermi surface properties from wannier interpolation," Phys. Rev. B, vol. 75, p. 195121, May 2007.

[11] M. V. Berry, "Quantal phase factors accompanying adiabatic changes," Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences, vol. 392, no. 1802, pp. 45–57, 1984.

[12] D. Xiao, M.-C. Chang, and Q. Niu, "Berry phase effects on electronic properties," Reviews of modern physics, vol. 82, no. 3, p. 1959, 2010.

[13] J. Zak, "Berry's phase for energy bands in solids," Phys. Rev. Lett., vol. 62, pp. 2747–2750, Jun 1989.

[14] R. King-Smith and D. Vanderbilt, "Theory of polarization of crystalline solids," vol. 47, pp. 1651–1654, 1993.

[15] D. Vanderbilt and R. King-Smith, "Electric polarization as a bulk quantity and its relation to surface charge," vol. 48, pp. 4442–4455, 1993.

[16] N. Nagaosa, J. Sinova, S. Onoda, A. MacDonald, and N. Ong, "Anomalous hall effect," Reviews of modern physics, vol. 82, no. 2, p. 1539, 2010.

[17] F. D. M. Haldane, "Berry curvature on the fermi surface: Anomalous hall effect as a topological fermi-liquid property," Phys. Rev. Lett., vol. 93, p. 206602, Nov 2004.

[18] X. Wang, J. R. Yates, I. Souza, and D. Vanderbilt, "*Ab initio* calculation of the anomalous hall conductivity by wannier interpolation," Phys. Rev. B, vol. 74, p. 195118, Nov 2006.

[19] X.-L. Qi, T. L. Hughes, and S.-C. Zhang, "Topological field theory of time-reversal invariant insulators," Phys. Rev. B, vol. 78, p. 195424, Nov 2008.

[20] A. M. Essin, J. E. Moore, and D. Vanderbilt, "Magnetoelectric polarizability and axion electrodynamics in crystalline insulators," Phys. Rev. Lett., vol. 102, p. 146805, Apr 2009.

[21] K. v. Klitzing, G. Dorda, and M. Pepper, "New method for high-accuracy determination of the fine-structure constant based on quantized hall resistance," Phys. Rev. Lett., vol. 45, pp. 494–497, Aug 1980.

[22] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, "Quantized hall conductance in a two-dimensional periodic potential," Phys. Rev. Lett., vol. 49, pp. 405–408, Aug 1982.

[23] M. Kohmoto, "Topological invariant and the quantization of the hall conductance," Annals of Physics, vol. 160, no. 2, pp. 343–354, 1985.

[24] F. D. M. Haldane, "Model for a quantum hall effect without landau levels: Condensed-matter realization of the "parity anomaly"," Phys. Rev. Lett., vol. 61, pp. 2015–2018, Oct 1988.

[25] C.-Z. Chang, J. Zhang, X. Feng, J. Shen, Z. Zhang, M. Guo, K. Li, Y. Ou, P. Wei, L.-L. Wang, et al., "Experimental observation of the quantum anomalous hall effect in a magnetic topological insulator," Science, vol. 340, no. 6129, pp. 167–170, 2013.

[26] C. L. Kane and E. J. Mele, "Quantum spin Hall effect in graphene," Phys. Rev. Lett., vol. 95, p. 226801, 2005.

[27] C. L. Kane and E. J. Mele, "Z_2 topological order and the quantum spin Hall effect," Phys. Rev. Lett., vol. 95, p. 146802, 2005.

[28] B. A. Bernevig, T. L. Hughes, and S. C. Zhang, "Quantum spin Hall effect and topological phase transition in HgTe quantum wells," Science, vol. 314, pp. 1757–1761, 2006.

[29] C. Wu, B. A. Bernevig, and S.-C. Zhang, "Helical liquid and the edge of quantum spin hall systems," Phys. Rev. Lett., vol. 96, p. 106401, Mar 2006.

[30] L. Fu, C. L. Kane, and E. J. Mele, "Topological insulators in three dimensions," Phys. Rev. Lett., vol. 98, p. 106803, 2007.

[31] J. E. Moore and L. Balents, "Topological invariants of time-reversal-invariant band structures," Physical Review B, vol. 75, no. 12, p. 121306, 2007.

[32] R. Roy, "Topological phases and the quantum spin hall effect in three dimensions," Physical Review B, vol. 79, no. 19, p. 195322, 2009.

[33] M. Z. Hasan and C. L. Kane, "*Colloquium* : Topological insulators," Rev. Mod. Phys., vol. 82, pp. 3045–3067, Nov 2010.

[34] X.-L. Qi and S.-C. Zhang, "Topological insulators and superconductors," Rev. Mod. Phys., vol. 83, pp. 1057–1110, Oct 2011.

[35] B. Yan and S.-C. Zhang, "Topological materials," Reports on Progress in Physics, vol. 75, no. 9, p. 096501, 2012.

[36] H. Zhang, C.-X. Liu, X.-L. Qi, X. Dai, Z. Fang, and S.-C. Zhang, "Topological insulators in Bi_2Se_3, Bi_2Te_3, and Sb_2Te_3 with a single Dirac cone on the surface," Nature Physics, vol. 5, pp. 438–442, 2009.

[37] Y. Xia, D. Qian, D. Hsieh, L. Wray, A. Pal, H. Lin, A. Bansil, D. Grauer, Y. Hor, R. Cava, et al., "Observation of a large-gap topological-insulator class with a single dirac cone on the surface," Nature Physics, vol. 5, no. 6, pp. 398–402, 2009.

[38] D. Hsieh, Y. Xia, D. Qian, L. Wray, F. Meier, J. H. Dil, J. Osterwalder, L. Patthey, A. V. Fedorov, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan, "Observation of time-reversal-protected single-dirac-cone topological-insulator states in bi_2te_3 and sb_2te_3," Phys. Rev. Lett., vol. 103, p. 146401, Sep 2009.

[39] Y. L. Chen, J. G. Analytis, J.-H. Chu, Z. K. Liu, S.-K. Mo, X. L. Qi, H. J. Zhang, D. H. Lu, X. Dai, Z. Fang, S. C. Zhang, I. R. Fisher, Z. Hussain, and Z.-X. Shen, "Experimental realization of a three-dimensional topological insulator, Bi_2Te_3," Science, vol. 325, no. 5937, pp. 178–181, 2009.

[40] L. Wu, M. Brahlek, R. V. Aguilar, A. Stier, C. Morris, Y. Lubashevsky, L. Bilbro, N. Bansal, S. Oh, and N. Armitage, "A sudden collapse in the transport lifetime across the topological phase transition in (bi1-xinx) 2se3," Nature Physics, vol. 9, pp. 410–414, 2013.

[41] M. Brahlek, N. Bansal, N. Koirala, S.-Y. Xu, M. Neupane, C. Liu, M. Z. Hasan, and S. Oh, "Topological-metal to band-insulator transition in $(bi_{1-x}in_x)_2se_3$ thin films," Phys. Rev. Lett., vol. 109, p. 186403, Oct 2012.

[42] "Topological phase transitions in $(bi_{1-x}in_x)_2se_3$ and $(bi_{1-x}sb_x)_2se_3$, author = Liu, Jianpeng and Vanderbilt, David, journal = Phys. Rev. B, volume = 88, issue = 22, pages = 224202, numpages = 12, year = 2013, month = Dec, publisher = American Physical Society, doi = 10.1103/PhysRevB.88.224202,,"

[43] X. Wan, A. M. Turner, A. Vishwanath, and S. Y. Savrasov, "Topological semimetal and fermi-arc surface states in the electronic structure of pyrochlore iridates," Phys. Rev. B, vol. 83, p. 205101, May 2011.

[44] A. A. Burkov and L. Balents, "Weyl semimetal in a topological insulator multilayer," Phys. Rev. Lett., vol. 107, p. 127205, Sep 2011.

[45] A. M. Turner, A. Vishwanath, and C. O. Head, "Beyond band insulators: Topology of semi-metals and interacting phases," Topological Insulators, vol. 6, p. 293, 2013.

[46] S. Murakami, "Phase transition between the quantum spin hall and insulator phases in 3d: emergence of a topological gapless phase," New Journal of Physics, vol. 9, no. 9, p. 356, 2007.

[47] S. Murakami and S.-i. Kuga, "Universal phase diagrams for the quantum spin hall systems," Phys. Rev. B, vol. 78, p. 165313, Oct 2008.

[48] J. Liu and D. Vanderbilt, "Weyl semimetals from noncentrosymmetric topological insulators," Phys. Rev. B, vol. 90, p. 155316, Oct 2014.

[49] S. Raghu, X.-L. Qi, C. Honerkamp, and S.-C. Zhang, "Topological mott insulators," Phys. Rev. Lett., vol. 100, p. 156401, Apr 2008.

[50] C. Wu and S.-C. Zhang, "Dynamic generation of spin-orbit coupling," Phys. Rev. Lett., vol. 93, p. 036403, Jul 2004.

[51] J. Liu and D. Vanderbilt, "Spin-orbit spillage as a measure of band inversion in insulators," Phys. Rev. B, vol. 90, p. 125133, Sep 2014.

[52] N. Marzari and D. Vanderbilt, "Maximally localized generalized Wannier functions for composite energy bands," Phys. Rev. B, vol. 56, pp. 12847–12865, 1997.

[53] I. Souza, N. Marzari, and D. Vanderbilt, "Maximally localized Wannier functions for entangled energy bands," Phys. Rev. B, vol. 65, p. 035109, 2001.

[54] R. Resta and D. Vanderbilt, "Theory of polarization: A modern approach," in Physics of Ferroelectrics: a Modern Perspective (C. H. A. K. M. Rabe and J.-M. Triscone, eds.), Berlin: Springer-Verlag, 2007.

[55] M. Tinkham, Group theory and quantum mechanics. Mineola, New York: Dover Publication, Inc., 2003.

[56] L. Fu and C. L. Kane, "Time reversal polarization and a Z_2 adiabatic spin pump," Phys. Rev. B, vol. 74, p. 195312, 2006.

[57] A. A. Soluyanov and D. Vanderbilt, "Smooth gauge for topological insulators," Phys. Rev. B, vol. 85, p. 115415, Mar 2012.

[58] T. Koga, J. Nitta, T. Akazaki, and H. Takayanagi, "Rashba spin-orbit coupling probed by the weak antilocalization analysis in InAlAs/InGaAs/InAlAs quantum wells as a function of quantum well asymmetry," Phys. Rev. Lett., vol. 89, p. 046801, Jul 2002.

[59] G. P. Mikitik and Y. V. Sharlai, "Manifestation of berry's phase in metal physics," Phys. Rev. Lett., vol. 82, pp. 2147–2150, Mar 1999.

[60] K. Novoselov, A. K. Geim, S. Morozov, D. Jiang, M. Katsnelson, I. Grigorieva, S. Dubonos, and A. Firsov, "Two-dimensional gas of massless dirac fermions in graphene," nature, vol. 438, no. 7065, pp. 197–200, 2005.

[61] A. Bohm, A. Mostafazadeh, H. Koizumi, Q. Niu, and J. Zwanziger, The Geometric phase in quantum systems: foundations, mathematical concepts, and applications in molecular and condensed matter physics. Berlin: Springer-Verlag, 2003.

[62] G. Panati, "Triviality of bloch and bloch–dirac bundles," in Annales Henri Poincaré, vol. 8, pp. 995–1011, Springer, 2007.

[63] R. Resta, "Macroscopic polarization in crystalline dielectrics: the geometric phase approach," Rev. Mod. Phys., vol. 66, pp. 899–915, Jul 1994.

[64] N. A. Spaldin, "A beginner's guide to the modern theory of polarization," Journal of Solid State Chemistry, vol. 195, pp. 2–10, 2012.

[65] R. Resta, "Theory of the electric polarization in crystals," Ferroelectrics, vol. 136, no. 1, pp. 51–55, 1992.

[66] R. Karplus and J. M. Luttinger, "Hall effect in ferromagnetics," Phys. Rev., vol. 95, pp. 1154–1160, Sep 1954.

[67] G. Sundaram and Q. Niu, "Wave-packet dynamics in slowly perturbed crystals: Gradient corrections and berry-phase effects," Phys. Rev. B, vol. 59, pp. 14915–14925, Jun 1999.

[68] T. Jungwirth, Q. Niu, and A. H. MacDonald, "Anomalous hall effect in ferromagnetic semiconductors," Phys. Rev. Lett., vol. 88, p. 207208, May 2002.

[69] X. Wang, D. Vanderbilt, J. R. Yates, and I. Souza, "Fermi-surface calculation of the anomalous hall conductivity," Phys. Rev. B, vol. 76, p. 195109, Nov 2007.

[70] M. Fiebig Journal of Physics D: Applied Physics, vol. 38, no. 8, p. R123, 2005.

[71] A. Malashevich, I. Souza, S. Coh, and D. Vanderbilt New Journal of Physics, vol. 12, no. 5, p. 053032, 2010.

[72] J. Íñiguez, "First-principles approach to lattice-mediated magnetoelectric effects," Phys. Rev. Lett., vol. 101, p. 117201, Sep 2008.

[73] E. Bousquet, N. A. Spaldin, and K. T. Delaney, "Unexpectedly large electronic contribution to linear magnetoelectricity," Phys. Rev. Lett., vol. 106, p. 107202, Mar 2011.

[74] A. M. Essin, A. M. Turner, J. E. Moore, and D. Vanderbilt, "Orbital magnetoelectric coupling in band insulators," Phys. Rev. B, vol. 81, p. 205104, May 2010.

[75] F. Wilczek, "Two applications of axion electrodynamics," Phys. Rev. Lett., vol. 58, pp. 1799–1802, May 1987.

[76] A. M. Turner, Y. Zhang, R. S. K. Mong, and A. Vishwanath, "Quantized response and topology of magnetic insulators with inversion symmetry," Phys. Rev. B, vol. 85, p. 165120, Apr 2012.

[77] D. Thouless, "Localisation and the two-dimensional hall effect," Journal of Physics C: Solid State Physics, vol. 14, no. 23, p. 3475, 1981.

[78] R. E. Prange, "Quantized hall resistance and the measurement of the fine-structure constant," Phys. Rev. B, vol. 23, pp. 4802–4805, May 1981.

[79] R. B. Laughlin, "Quantized hall conductivity in two dimensions," Phys. Rev. B, vol. 23, pp. 5632–5633, May 1981.

[80] R. B. Laughlin, "Nobel lecture: Fractional quantization," Rev. Mod. Phys., vol. 71, pp. 863–874, Jul 1999.

[81] B. I. Halperin, "Quantized hall conductance, current-carrying edge states, and the existence of extended states in a two-dimensional disordered potential," Phys. Rev. B, vol. 25, pp. 2185–2190, Feb 1982.

[82] T. Thonhauser and D. Vanderbilt, "Insulator/chern-insulator transition in the haldane model," Phys. Rev. B, vol. 74, p. 235111, Dec 2006.

[83] C.-X. Liu, X.-L. Qi, X. Dai, Z. Fang, and S.-C. Zhang, "Quantum anomalous hall effect in $hg_{1-y}mn_yTe$ quantum wells," Phys. Rev. Lett., vol. 101, p. 146802, Oct 2008.

[84] Z. Qiao, S. A. Yang, W. Feng, W.-K. Tse, J. Ding, Y. Yao, J. Wang, and Q. Niu, "Quantum anomalous hall effect in graphene from rashba and exchange effects," Phys. Rev. B, vol. 82, p. 161414, Oct 2010.

[85] R. Yu, W. Zhang, H.-J. Zhang, S.-C. Zhang, X. Dai, and Z. Fang, "Quantized anomalous hall effect in magnetic topological insulators," Science, vol. 329, no. 5987, pp. 61–64, 2010.

[86] H. Zhang, C. Lazo, S. Blügel, S. Heinze, and Y. Mokrousov, "Electrically tunable quantum anomalous hall effect in graphene decorated by $5d$ transition-metal adatoms," Phys. Rev. Lett., vol. 108, p. 056802, Feb 2012.

[87] K. F. Garrity and D. Vanderbilt, "Chern insulators from heavy atoms on magnetic substrates," Phys. Rev. Lett., vol. 110, p. 116802, Mar 2013.

[88] S. Datta, Electronic transport in mesoscopic systems. Cambridge university press, 1997.

[89] M. Konig, S. Wiedmann, C. Brune, A. Roth, H. Buhmann, L. W. Molenkamp, X. L. Qi, and S. C. Zhang, "Quantum spin hall insulator state in HgTe quantum wells," Science, vol. 318, pp. 766–770, 2007.

[90] L. Fu and C. L. Kane, "Topological insulators with inversion symmetry," Phys. Rev. B, vol. 76, p. 045302, 2007.

[91] S. Murakami, "Quantum spin hall effect and enhanced magnetic response by spin-orbit coupling," Phys. Rev. Lett., vol. 97, p. 236805, Dec 2006.

[92] C.-C. Liu, W. Feng, and Y. Yao, "Quantum spin hall effect in silicene and two-dimensional germanium," Phys. Rev. Lett., vol. 107, p. 076802, Aug 2011.

[93] A. Shitade, H. Katsura, J. Kuneš, X.-L. Qi, S.-C. Zhang, and N. Nagaosa, "Quantum spin hall effect in a transition metal oxide na$_2$iro$_3$," Phys. Rev. Lett., vol. 102, p. 256403, Jun 2009.

[94] Y. Xu, B. Yan, H.-J. Zhang, J. Wang, G. Xu, P. Tang, W. Duan, and S.-C. Zhang, "Large-gap quantum spin hall insulators in tin films," Phys. Rev. Lett., vol. 111, p. 136804, Sep 2013.

[95] S. Coh, D. Vanderbilt, A. Malashevich, and I. Souza, "Chern-simons orbital magnetoelectric coupling in generic insulators," Phys. Rev. B, vol. 83, p. 085108, Feb 2011.

[96] H. Lin, L. A. Wray, Y. Xia, S. Xu, S. Jia, R. J. Cava, A. Bansil, and M. Z. Hasan, "Half-heusler ternary compounds as new multifunctional experimental platforms for topological quantum phenomena," Nature Materials, vol. 9, no. 7, pp. 546–549, 2010.

[97] H. Lin, R. S. Markiewicz, L. A. Wray, L. Fu, M. Z. Hasan, and A. Bansil, "Single-dirac-cone topological surface states in the tlbise$_2$ class of topological semiconductors," Phys. Rev. Lett., vol. 105, p. 036404, Jul 2010.

[98] T. Sato, K. Segawa, H. Guo, K. Sugawara, S. Souma, T. Takahashi, and Y. Ando, "Direct evidence for the dirac-cone topological surface states in the ternary chalcogenide tlbise$_2$," Phys. Rev. Lett., vol. 105, p. 136802, Sep 2010.

[99] K. Kuroda, M. Ye, A. Kimura, S. V. Eremeev, E. E. Krasovskii, E. V. Chulkov, Y. Ueda, K. Miyamoto, T. Okuda, K. Shimada, H. Namatame, and M. Taniguchi, "Experimental realization of a three-dimensional topological insulator phase in ternary chalcogenide tlbise$_2$," Phys. Rev. Lett., vol. 105, p. 146801, Sep 2010.

[100] J. Chen, H. J. Qin, F. Yang, J. Liu, T. Guan, F. M. Qu, G. H. Zhang, J. R. Shi, X. C. Xie, C. L. Yang, K. H. Wu, Y. Q. Li, and L. Lu, "Gate-voltage control of chemical potential and weak antilocalization in bi$_2$se$_3$," Phys. Rev. Lett., vol. 105, p. 176602, Oct 2010.

[101] J. G. Checkelsky, Y. S. Hor, R. J. Cava, and N. P. Ong, "Bulk band gap and surface state conduction observed in voltage-tuned crystals of the topological insulator bi$_2$se$_3$," Phys. Rev. Lett., vol. 106, p. 196801, May 2011.

[102] D. Kong, Y. Chen, J. J. Cha, Q. Zhang, J. G. Analytis, K. Lai, Z. Liu, S. S. Hong, K. J. Koski, S.-K. Mo, et al., "Ambipolar field effect in the ternary topological insulator (bixsb1-x) 2te3 by composition tuning," Nature nanotechnology, vol. 6, no. 11, pp. 705–709, 2011.

[103] A. A. Taskin, Z. Ren, S. Sasaki, K. Segawa, and Y. Ando, "Observation of dirac holes and electrons in a topological insulator," Phys. Rev. Lett., vol. 107, p. 016801, Jun 2011.

[104] L. Fu and C. L. Kane, "Superconducting proximity effect and majorana fermions at the surface of a topological insulator," Phys. Rev. Lett., vol. 100, p. 096407, Mar 2008.

[105] Q. D. Gibson, L. M. Schoop, L. Muechler, L. S. Xie, M. Hirschberger, N. P. Ong, R. Car, and R. J. Cava, "Three-dimensional dirac semimetals: Design principles and predictions of new materials," Phys. Rev. B, vol. 91, p. 205128, May 2015.

[106] J. E. Moore and L. Balents, "Topological invariants of time-reversal-invariant band structures," Phys. Rev. B, vol. 75, p. 121306, 2007.

[107] C.-X. Liu, X.-L. Qi, H. Zhang, X. Dai, Z. Fang, and S.-C. Zhang, "Model hamiltonian for topological insulators," Phys. Rev. B, vol. 82, p. 045122, Jul 2010.

[108] D. Hsieh, Y. Xia, L. Wray, D. Qian, A. Pal, J. Dil, J. Osterwalder, F. Meier, G. Bihlmayer, C. Kane, et al., "Observation of unconventional quantum spin textures in topological insulators," Science, vol. 323, no. 5916, pp. 919–922, 2009.

[109] L. Wray, S. Xu, Y. Xia, D. Hsieh, A. Fedorov, Y. San Hor, R. Cava, A. Bansil, H. Lin, and M. Hasan, "A topological insulator surface under strong coulomb, magnetic and disorder perturbations," Nature Physics, vol. 7, no. 1, pp. 32–37, 2010.

[110] D. Hsieh, Y. Xia, D. Qian, L. Wray, J. Dil, F. Meier, J. Osterwalder, L. Patthey, J. Checkelsky, N. Ong, et al., "A tunable topological insulator in the spin helical dirac transport regime," Nature, vol. 460, no. 7259, pp. 1101–1105, 2009.

[111] W. Feng and Y. Yao, "Three-dimensional topological insulators: A review on host materials," Science China Physics, Mechanics and Astronomy, vol. 55, no. 12, pp. 2199–2212, 2012.

[112] T. Sato, K. Segawa, K. Kosaka, S. Souma, K. Nakayama, K. Eto, T. Minami, Y. Ando, and T. Takahashi, "Unexpected mass acquisition of dirac fermions at the quantum phase transition of a topological insulator," Nature Physics, vol. 7, no. 11, pp. 840–844, 2011.

[113] S. Xu, Y. Xia, L. Wray, S. Jia, F. Meier, J. Osterwalder, B. Slomski, A. Bansil, H. Lin, R. Cava, et al., "Topological phase transition and texture inversion in a tunable topological insulator," Science, vol. 332, no. 6029, pp. 560–564, 2011.

[114] S. Chadov, J. Kiss, C. Felser, K. Chadova, D. Ködderitzsch, J. Minár, and H. Ebert, "Topological phase transition induced by random substitution," arXiv preprint arXiv:1207.3463, 2012.

[115] W. Ku, T. Berlijn, and C.-C. Lee, "Unfolding first-principles band structures," Phys. Rev. Lett., vol. 104, p. 216401, May 2010.

[116] T. Berlijn, D. Volja, and W. Ku, "Can disorder alone destroy the e'_g hole pockets of Na_xCoO_2? a wannier function based first-principles method for disordered systems," Phys. Rev. Lett., vol. 106, p. 077005, Feb 2011.

[117] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch, "Quantum espresso: a modular and open-source software project for quantum simulations of materials," Journal of Physics: Condensed Matter, vol. 21, no. 39, p. 395502 (19pp), 2009.

[118] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," Phys. Rev. Lett., vol. 77, pp. 3865–3868, Oct 1996.

[119] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple [phys. rev. lett. 77, 3865 (1996)]," Phys. Rev. Lett., vol. 78, pp. 1396–1396, Feb 1997.

[120] D. Vanderbilt, "Soft self-consistent pseudopotentials in a generalized eigenvalue formalism," vol. 41, pp. 7892–7895, 1990.

[121] http://opium.sourceforge.net/.

[122] N. J. Ramer and A. M. Rappe, "Designed nonlocal pseudopotentials for enhanced transferability," Phys. Rev. B, vol. 59, pp. 12471–12478, May 1999.

[123] H. J. Monkhorst and J. D. Pack, "Special points for brillouin-zone integrations," Phys. Rev. B, vol. 13, pp. 5188–5192, Jun 1976.

[124] A. A. Mostofi, J. R. Yates, Y. S. Lee, I. Souza, D. Vanderbilt, and N. Marzari, "Wannier90: A tool for obtaining maximally-localised Wannier functions," Computer Phys. Comm., vol. 178, pp. 685–699, 2008.

[125] D. Eddike, A. Ramdani, G. Brun, J. Tedenac, and B. Liautard, "Phase diagram equilibria in2se3-sb2se3 crystal growth of the [beta]-in2se3 phase (in1. 94sb0. 06se3)," Materials research bulletin, vol. 33, no. 3, pp. 519–523, 1998.

[126] X. Zhu and S. G. Louie, "Quasiparticle band structure of thirteen semiconductors and insulators," Phys. Rev. B, vol. 43, pp. 14142–14156, Jun 1991. The CBM of InAs at Γ is dominated by In 5s orbitals.

[127] J. P. Perdew and A. Zunger, "Self-interaction correction to density-functional approximations for many-electron systems," Phys. Rev. B, vol. 23, pp. 5048–5079, May 1981.

[128] http://www.wien2k.at/.

[129] G. Kresse and J. Furthmüller, "Software vasp, vienna (1999)," Phys. Rev. B, vol. 54, no. 11, p. 169, 1996.

[130] A. A. Soluyanov and D. Vanderbilt, "Computing topological invariants without inversion symmetry," Phys. Rev. B, vol. 83, p. 235401, JUN 2 2011.

[131] V. Popescu and A. Zunger, "Effective band structure of random alloys," Phys. Rev. Lett., vol. 104, p. 236403, Jun 2010.

[132] B. Leung and E. Prodan, "Effect of strong disorder in a three-dimensional topological insulator: Phase diagram and maps of the $\mathcal{F}_2$ invariant," Phys. Rev. B, vol. 85, p. 205136, May 2012.

[133] J. Bellissard, A. van Elst, and H. Schulz-Baldes, "The noncommutative geometry of the quantum hall effect," Journal of Mathematical Physics, vol. 35, pp. 5373–5451, 1994.

[134] E. Prodan, "Disordered topological insulators: a non-commutative geometry perspective," Journal of Physics A: Mathematical and Theoretical, vol. 44, no. 11, p. 113001, 2011.

[135] H. Schulz-Baldes and S. Teufel, "Orbital polarization and magnetization for independent particles in disordered media," Communications in Mathematical Physics, vol. 319, p. 649, 2013.

[136] B. Leung and E. Prodan, "A non-commutative formula for the isotropic magnetoelectric response," Journal of Physics A: Mathematical and Theoretical, vol. 46, no. 8, p. 085205, 2013.

[137] J. Henk, M. Flieger, I. V. Maznichenko, I. Mertig, A. Ernst, S. V. Eremeev, and E. V. Chulkov, "Topological character and magnetism of the dirac state in mn-doped bi_2te_3," Phys. Rev. Lett., vol. 109, p. 076801, Aug 2012.

[138] J. Henk, A. Ernst, S. V. Eremeev, E. V. Chulkov, I. V. Maznichenko, and I. Mertig, "Complex spin texture in the pure and mn-doped topological insulator bi_2te_3," Phys. Rev. Lett., vol. 108, p. 206801, May 2012.

[139] H. B. Nielsen and M. Ninomiya, "The adler-bell-jackiw anomaly and weyl fermions in a crystal," Physics Letters B, vol. 130, no. 6, pp. 389–396, 1983.

[140] A. A. Burkov, M. D. Hook, and L. Balents, "Topological nodal semimetals," Phys. Rev. B, vol. 84, p. 235126, Dec 2011.

[141] P. Hosur and X. Qi, "Recent developments in transport phenomena in weyl semimetals," Comptes Rendus Physique, vol. 14, no. 9, pp. 857–870, 2013.

[142] R. Okugawa and S. Murakami, "Dispersion of fermi arcs in weyl semimetals and their evolutions to dirac cones," Phys. Rev. B, vol. 89, p. 235315, Jun 2014.

[143] K.-Y. Yang, Y.-M. Lu, and Y. Ran, "Quantum hall effects in a weyl semimetal: Possible application in pyrochlore iridates," Phys. Rev. B, vol. 84, p. 075129, Aug 2011.

[144] F. Haldane, "Attachment of surface" fermi arcs" to the bulk fermi surface:" fermi-level plumbing" in topological metals," arXiv preprint arXiv:1401.0529, 2014.

[145] A. A. Zyuzin and A. A. Burkov, "Topological response in weyl semimetals and the chiral anomaly," Phys. Rev. B, vol. 86, p. 115133, Sep 2012.

[146] V. Aji, "Adler-bell-jackiw anomaly in weyl semimetals: Application to pyrochlore iridates," Phys. Rev. B, vol. 85, p. 241101, Jun 2012.

[147] D. Bulmash, C.-X. Liu, and X.-L. Qi, "Prediction of a weyl semimetal in $hg_{1-x-y}cd_xmn_yte$," Phys. Rev. B, vol. 89, p. 081106, Feb 2014.

[148] B.-J. Yang, M. S. Bahramy, R. Arita, H. Isobe, E.-G. Moon, and N. Nagaosa, "Theory of topological quantum phase transitions in 3d noncentrosymmetric systems," Phys. Rev. Lett., vol. 110, p. 086402, Feb 2013.

[149] M. Bahramy, B.-J. Yang, R. Arita, and N. Nagaosa, "Emergence of noncentrosymmetric topological insulating phase in bitei under pressure," Nature communications, vol. 3, p. 679, 2012.

[150] B. Yan, H.-J. Zhang, C.-X. Liu, X.-L. Qi, T. Frauenheim, and S.-C. Zhang, "Theoretical prediction of topological insulator in ternary rare earth chalcogenides," Phys. Rev. B, vol. 82, p. 161108, Oct 2010.

[151] X. Xi, C. Ma, Z. Liu, Z. Chen, W. Ku, H. Berger, C. Martin, D. B. Tanner, and G. L. Carr, "Signatures of a pressure-induced topological quantum phase transition in bitei," Phys. Rev. Lett., vol. 111, p. 155701, Oct 2013.

[152] K. Ishizaka, M. Bahramy, H. Murakawa, M. Sakano, T. Shimojima, T. Sonobe, K. Koizumi, S. Shin, H. Miyahara, A. Kimura, K. Miyamoto, T. Okuda, H. Namatame, M. Taniguchi, R. Arita, N. Nagaosa, Y. Kobayashi, K Murakami, R. Kumai, Y. Kaneko, Y. Onose, and Y. Tokura, "Giant rashba-type spin splitting in bulk bitei," Nature materials, vol. 10, no. 7, pp. 521–526, 2011.

[153] G. Kresse and J. Furthmüller, "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set," Phys. Rev. B, vol. 54, pp. 11169–11186, Oct 1996.

[154] G. Kresse and J. Furthmüller, "Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set," Computational Materials Science, vol. 6, no. 1, pp. 15–50, 1996.

[155] R. Winkler, Spin-orbit coupling effects in two-dimensional electron and hole systems, vol. 191. Springer Verlag, 2003.

[156] B. Nikolić, L. Zarbo, and S. Souma, The Oxford Handbook on Nanoscience and Technology: Frontiers and Advances. Oxford University Press, 2010. edited by A. V. Narlikar and Y. Y. Fu, Vol. I, Chap. 24, pp. 814866.

[157] S. Hikami, A. I. Larkin, and Y. Nagaoka, "Spin-orbit interaction and magnetoresistance in the two dimensional random system," Progress of Theoretical Physics, vol. 63, no. 2, pp. 707–710, 1980.

[158] J. E. Hirsch, "Spin hall effect," Phys. Rev. Lett., vol. 83, pp. 1834–1837, Aug 1999.

[159] H. Zhang, F. Freimuth, G. Bihlmayer, M. Ležaić, S. Blügel, and Y. Mokrousov, "Engineering quantum anomalous hall phases with orbital and spin degrees of freedom," Phys. Rev. B, vol. 87, p. 205132, May 2013.

[160] M. Klintenberg, "The search for strong topological insulators," arXiv:1007.4838, 2010.

[161] K. Yang, W. Setyawan, S. Wang, M. B. Nardelli, and S. Curtarolo, "A search model for topological insulators with high-throughput robustness descriptors," Nature Materials, vol. 11, no. 7, pp. 614–619, 2012.

[162] H. Zhang, F. Freimuth, G. Bihlmayer, S. Blügel, and Y. Mokrousov, "Topological phases of bi(111) bilayer in an external exchange field," Phys. Rev. B, vol. 86, p. 035104, Jul 2012.

[163] Y. Liu and R. E. Allen, "Electronic structure of the semimetals bi and sb," Phys. Rev. B, vol. 52, pp. 1566–1577, Jul 1995.

[164] T. Fukui, Y. Hatsugai, and H. Suzuki, "Chern numbers in discretized brillouin zone: Efficient method of computing (spin) hall conductances," Journal of the Physical Society of Japan, vol. 74, no. 6, pp. 1674–1677, 2005.

[165] T. L. Hughes, E. Prodan, and B. A. Bernevig, "Inversion-symmetric topological insulators," Phys. Rev. B, vol. 83, p. 245132, Jun 2011.

[166] X. L. Qi, T. L. Hughes, and S. C. Zhang, "Topological field theory of time-reversal invariant insulators," Phys. Rev. B, vol. 78, p. 195424, 2008.

[167] R. Yu, X. L. Qi, A. Bernevig, Z. Fang, and X. Dai, "Equivalent expression of F_2 topological invariant for band insulators using the non-abelian berry connection," Phys. Rev. B, vol. 84, p. 075119, Aug 2011.

[168] M. Taherinejad and D. Vanderbilt, "Adiabatic pumping of chern-simons axion coupling," Phys. Rev. Lett., vol. 114, p. 096401, Mar 2015.

[169] J. Li, R.-L. Chu, J. K. Jain, and S.-Q. Shen, "Topological anderson insulator," Phys. Rev. Lett., vol. 102, p. 136806, Apr 2009.

[170] D. Pesin and L. Balents, "Mott physics and band topology in materials with strong spin–orbit interaction," Nature Physics, vol. 6, no. 5, pp. 376–381, 2010.

[171] M. Dzero, K. Sun, V. Galitski, and P. Coleman, "Topological kondo insulators," Phys. Rev. Lett., vol. 104, p. 106408, Mar 2010.

[172] Z. Wang and S.-C. Zhang, "Simplified topological invariants for interacting insulators," Phys. Rev. X, vol. 2, p. 031008, Aug 2012.

[173] Z. Huang and D. P. Arovas, "Topological indices for open and thermal systems via uhlmann's phase," Phys. Rev. Lett., vol. 113, p. 076407, Aug 2014.

[174] O. Viyuela, A. Rivas, and M. A. Martin-Delgado, "Two-dimensional density-matrix topological fermionic phases: Topological uhlmann numbers," Phys. Rev. Lett., vol. 113, p. 076408, Aug 2014.

[175] R. Lou, Z. Liu, W. Jin, H. Wang, Z. Han, K. Liu, X. Wang, T. Qian, Y. Kushnirenko, S.-W. Cheong, R. M. Osgood Jr., H. Ding, and S. Wang, "A sudden gap-closure cross the topological phase transition in $(\mathrm{bi}_x\mathrm{in}_{1-x})_2\mathrm{se}_3$ single crystals," arXiv preprint arXiv:1503.07674, 2015.

[176] A. Rüegg and G. A. Fiete, "Topological insulators from complex orbital order in transition-metal oxides heterostructures," Phys. Rev. B, vol. 84, p. 201103, Nov 2011.

[177] W. Witczak-Krempa, G. Chen, Y. B. Kim, and L. Balents, "Correlated quantum phenomena in the strong spin-orbit regime," Annu. Rev. Condens. Matter Phys, vol. 5, p. 57, 2013.