TOPOLOGICAL ASPECTS OF BAND THEORY

BY ALEXEY A. SOLUYANOV

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, 2012

ABSTRACT OF THE DISSERTATION

Topological aspects of band theory

by Alexey A. Soluyanov

Dissertation Director: David Vanderbilt

Band theory has proven to be one of the most successful developments in condensed matter theory. It is the basis of our current understanding of crystalline solids, describing complex electronic behavior in terms of a single quasi-particle that moves in some effective field of the crystal lattice environment and other particles. In recent years topological and geometrical considerations opened a fundamentally new branch of research in band theory. One of the major advances in this field came with the realization that insulating band structures can be classified according to the values of some topological invariants associated with the occupied single-particle states. Insulators that correspond to non-trivial values of these topological invariants realize new states of matter with properties drastically different from those attributed to an ordinary insulator.

In this work we address questions that arise in the context of band theory in the presence of topologically non-trivial bands. Part of the thesis is aimed at the actual determination of the presence of non-trivial band topology. We develop a method to distinguish an ordinary insulator from a topological one in the presence of time-reversal symmetry. The method is implemented within the density functional theory framework and is illustrated with applications to real materials in *ab initio* calculations.

Another question considered in this work is that of a real-space representation of topological insulators, and in particular, the construction of Wannier functions – localized real-space wavefunctions. Wannier functions form one of the most powerful tools in band theory, and it is very important to understand how to implement Wannier function techniques in the presence of topological bands. In some cases bands with non-trivial topology do not allow for the construction of exponentially localized Wannier functions. While previous work has shown that in the presence of time-reversal symmetry such a construction should be possible in principle, it has remained unclear how to do it in practice.

We present an explicit construction of a Wannier representation for a particular model of a time-reversal invariant topological insulator. This construction is very different from the one used for ordinary band insulators. We then proceed to develop a procedure that allows for such a construction in the general case, without any reference to a particular model. Our work provides a basis for extending Wannier function techniques to topologically non-trivial band structures.

Acknowledgements

This work would not be possible if not for the wise supervision of my advisor David Vanderbilt. From him I learned most of the physics I know, and inspiration that came along with our discussions was filling my sails with wind whenever I was stuck on something. Being David's student I always enjoyed freedom to choose a problem of interest, which I consider to be the most wonderful privilege for a student. But it is not only the opportunities in physics that I would like to thank David for – his support during some hard times of my life helped me a lot in leaving it behind.

Our condensed matter group at Rutgers was always very helpful and I would like to thank Karin Rabe, Donald Hamann, Morrel Cohen, Natan Andrei, Lev Ioffe, Emil Yuzbashyan, Girsh Blumberg, Paul Leath, Victor Oudovenko, Gabriel Kotliar, Kristjan Haule and David Langreth for being so open whenever I had a question. A special word of gratitude should be said to Piers Coleman, who not only shared his excitement discussing physics with me, but spent a lot of time being one of my references, and to Premi Chandra for giving me one very valuable piece of advice. I would also like to acknowledge Emil Prodan of Yeshiva University for helpful discussions and for his patience with my recommendation letters.

During my five years at Rutgers I was lucky to be always surrounded by a community of bright students and postdocs. Discussions with Sinisa Coh, Anindya Roy, Maryam Taheri-Nejad, Jianpeng Liu, Andrei Malashevich, Xifan Wu, Aline Ramires, Victor Alexandrov, Tahir Yusufaly, Chuck-Hou Yee, Evgeniy Andriyash, Kostyantin Kechedzhi, Reza Nourafkan, Andriy Nevidomskyy,

Maxim Dzero, Éamonn Murray, Kevin Garrity, Joseph Bennett, Oscar Paz, Oswaldo Diéguez, Jiawang Hong, Kyuho Lee, Maxim Kharitonov, Bryan Leung, Qibin Zhou, Yangpeng Yao, Wenhu Xu and Matthew Foster were very fruitful. This list is far from being complete.

I would like to thank Ronald Ransome for his enormous help during these years as the Graduate Director and later as the department head, Mohan Kalelkar and Weida Wu for being my graduate committee members, and Gene Mele of University of Pennsylvania for kindly agreeing to be my external committee member.

All in all, Rutgers was very welcoming to me during all these years and I would like to thank all those people who were my teachers, who were my students, and who helped me with all the enormous amounts of paperwork.

Last but not least I thank my family Natalia, Alexander, Evguenia, Catherine, Irina and Gosha for their perpetual support and love.

Table of Contents

Abstract ii

Acknowledgements iv

List of Tables ix

List of Figures x

1. Introduction 1

2. Basic notions of band theory 8

2.1. From the many-body problem to many single-body problems . . . 8

2.1.1. Crystal potential and Bloch theorem 11

2.1.2. Reciprocal space, Bloch states, and band structure 16

2.2. Density functional theory 21

2.2.1. General formalism of DFT 21

2.2.2. Kohn-Sham solution scheme 22

2.2.3. Exchange-correlation potential 25

2.2.4. Pseudopotentials: simplifying atomic potentials 26

2.2.5. Inclusion of spin-orbit interaction 29

2.3. Tight-binding approximation 31

3. Topological insulators 36

3.1. Gauge transformations 38

3.2. Geometric phases in band theory 40

3.2.1. Berry phase 40

3.2.2. Applications in band theory 42

3.2.3. Numerical computation of Berry curvature 44

3.3. Chern insulators 46

3.4. $\mathbb{Z}_2$ insulators 50

3.4.1. Band structure in the presence of $\mathcal{T}$ symmetry 51

3.4.2. 2D: Quantum spin Hall phase 52

3.4.3. Kane-Mele model 55

3.4.4. Computing the $\mathbb{Z}_2$ invariant 59

3.4.5. 3D: surface metal 63

4. Wannier functions 66

4.1. Maximally localized Wannier functions 68

4.2. Hybrid Wannier functions 70

4.3. Chern numbers via hybrid WFs 71

4.4. Parallel transport gauge 74

5. Computing topological invariants without inversion symmetry 77

5.1. $\mathbb{Z}_2$ invariant via Wannier charge centers 79

5.1.1. Review of time reversal polarization 79

5.1.2. Formulation in terms of Wannier charge centers 82

5.2. Numerical Implementation 84

5.3. Numerical Implementation 84

5.3.1. Tracking WCC locations 85

5.3.2. Tracking gaps in the WCC spectrum 87

5.3.3. Application to 2D and 3D $\mathcal{T}$-invariant insulators 90

5.4. Application to real materials 91

5.4.1. Centrosymmetric materials 92

5.4.2. Noncentrosymmetric materials 95

6. Wannier functions for topological insulators 100

6.1. Explicit construction of Wannier functions 101

6.1.1. $\mathbb{Z}_2$-even phase 102

6.1.2. $\mathbb{Z}_2$-odd phase 105

6.2. Localization of Wannier Functions in the $\mathbb{Z}_2$-odd Insulator 108

6.3. Hybrid WCCs and polarization of Kane-Mele model 110

6.3.1. Total polarization 111

6.3.2. Hybrid Wannier decomposition 112

6.3.3. Results 113

6.4. Generalization to 3D 116

7. Smooth gauge for topological insulators 118

7.1. Cylindrical gauge and individual Chern numbers 121

7.1.1. Single band case 121

7.1.2. Multiband case and individual Chern numbers 123

7.2. Decomposition into Chern subspaces 124

7.2.1. Moving the gauge discontinuity to the BZ edge 125

7.2.2. Disentangling the two bands 130

7.2.3. Establishing a cylindrical gauge 135

7.2.4. Relation to spin Chern numbers 136

7.3. Rotation into a smooth gauge 138

7.4. Time-reversal constraint and smooth gauge 141

8. Summary and conclusions 144

Bibliography 146

Vita 160

List of Tables

3.1. Nonzero coefficients appearing in Eq. (3.34), using the notation $x = k_x a/2$ and $y = \sqrt{3}k_y a/2$ (see also Fig. 3.6). Adopted from Ref. [1]. 57

List of Figures

1.1. Possible band structures for a 1D system with two electrons per cell (lattice constant is taken to be a unit of length). Five lowest energy bands are plotted. (a) Metallic system. No energy gap. (b) Insulating system. The energy gap separates two occupied bands. 2

2.1. Reciprocal space (**k**-space). Brillouin zone (shown in blue) is formed by the primitive reciprocal lattice vectors $\mathbf{b}_1$ and $\mathbf{b}_2$. A non-primitive reciprocal lattice vector $\mathbf{G}_m$ is shown in green. 13

2.2. Energy band in 1D k-space. 17

2.3. BZ in 2D is a torus. 18

2.4. Energy levels of valence electrons (a) form a band structure in crystalline environment (b). Core states are lower in energy. . . . 20

3.1. Honeycomb lattice for Haldene and Kane-Mele models. 47

3.2. Schematic edge spectrum of the semi-infinite Chern insulator: (a) $C = 1$, (b) $C = -1$, (c) $C = 0$, (d) $C = 1$. The continuum of bulk bands is shaded. 49

3.3. Edge of a semi-infinite sample. Arrows indicate continuation to infinity. 49

3.4. Band structure for a hypothetical semi-infinite sample of the quantum spin Hall insulator. Shadowed regions correspond to the continuum of bulk states. 53

3.5. Gapping the edge spectrum by adiabatic changes in the Hamiltonian. 55

3.6. Brillouin zone sketched using coordinates $x = k_x a/2$ and $y = \sqrt{3}k_y a/2$. Primitive reciprocal lattice vectors $\mathbf{G}_1 = (2\pi/a)(1, 1/\sqrt{3})$ and $\mathbf{G}_2 = (2\pi/a)(-1, 1/\sqrt{3})$ correspond to $\mathbf{g}_1 = (\pi, \pi)$ and $\mathbf{g}_2 = (-\pi, \pi)$ respectively. The black rectangle marks the boundary $\partial\zeta$ of the zone used for polarization calculations in Sec. 6.3. 58

3.7. Phase diagram of the Kane-Mele model for $\lambda_{\rm v}/\lambda_{\rm SO} > 0$. Arrow illustrates a path crossing the phase boundary by varying $\lambda_{\rm v}$ while keeping other parameters fixed. 59

3.8. Sketch of the Brillouin zone. The Berry curvature of Eq. (3.38) is calculated in the interior of the half zone τ (dashed region), while the Berry connection is evaluated along its boundary $\partial\tau$ (arrows indicate direction of integration). Time-reversal–invariant points Γ_i are shown. 60

4.1. A sketch of amplitudes of Wannier functions in a 1D crystal. Solid curve refers to the home unit cell; dashed curves refer to periodic images in the neighboring cells. 67

4.2. A sketch of $\bar{x}(k_y)$ dependence. Panel (a): $C = 0$. Panel (b): $C = 1$. 72

5.1. Sketch of evolution of Wannier charge centers (WCCs) $\bar{x}$ vs. time t during an adiabatic pumping process. Regarding $\bar{x} \in [0, 1]$ as a unit circle and $t \in [0, T/2]$ as a line segment, the cylindrical $(\bar{x}, t)$ manifold is represented via a sequence of circular cross sections at left, or as an unwrapped cylinder at right. Each red rhombus marks the middle of the largest gap between WCCs at given t. (a) $\mathbb{Z}_2$ insulator; WCCs wind around the cylinder. (b) Normal insulator; WCCs reconnect without wrapping the cylinder. 85

5.2. Sketch illustrating the method used to determine whether $\bar{x}_n^{(m+1)}$ lies between $z^{(m)}$ and $z^{(m+1)}$ in the counterclockwise sense when mapped onto the complex unit circle. (a) Yes, since the directed area of the triangle is positive. (b) No, since it is negative. 89

5.3. Band structure of Bi along high symmetry lines in the BZ. Fermi level is shown in red to illustrate the semimetallic nature of Bi. . . 92

5.4. Band structure of $\mathrm{Bi_2Se_3}$ along high symmetry lines in the BZ. . . 92

5.5. Evolution of Bi WCCs $\bar{x}_n$ (circles) in the r_3 direction vs. k_2 at (a) k_1=0; (b) k_1=π. Red rhombus marks midpoint of largest gap. k_2 is sampled in ten equal increments from 0 to π, except that an extra point is inserted midway in the last segment in panel (b) (see text). 94

5.6. Evolution of $\mathrm{Bi_2Se_3}$ WCCs $\bar{x}_n$ (circles) in the r_3 direction vs. k_2 at (a) k_1=0; (b) k_1=π. Red rhombus marks midpoint of largest gap. k_2 is sampled in ten equal increments from 0 to π. 95

5.7. Band structure of GeTe along high symmetry lines in the BZ. . . 96

5.8. Evolution of GeTe WCCs $\bar{x}_n$ (circles) in the r_3 direction vs. k_2 at (a) k_1=0; (b) k_1=π. Red rhombus marks midpoint of largest gap. k_2 is sampled in ten equal increments from 0 to π. 97

5.9. Band structure along high-symmetry lines of the undistorted FCC structure for HgTe under tensile epitaxial strain in the [111] direction. (a) +2% epitaxial strain. (b) +5% epitaxial strain. 98

5.10. Evolution of WCCs for HgTe under +2% epitaxial strain in the [111] direction. WCCs $\bar{x}_n$ (circles) in the r_3 direction are plotted vs. k_2 at (a) k_1=0; (b) k_1=π. Red rhombus marks midpoint of largest gap. k_2 is sampled in ten equal increments from 0 to π, except that an extra point is inserted midway in the first segment in panel (a) (see text). 98

6.1. Sum of the weights of the projections into the two occupied bands of the basis states $|A;\uparrow_z\rangle$, $|B;\uparrow_z\rangle$, $|A;\downarrow_z\rangle$, and $|B;\downarrow_z\rangle$ plotted along the diagonal of the BZ for (a) $\lambda_v/t = 5$ ($\mathbb{Z}_2$-even phase) and (b) $\lambda_v/t = 1$ ($\mathbb{Z}_2$-odd phase). Inset in (a): BZ of a honeycomb lattice. 103

6.2. Plot of $\det[S(\mathbf{k})]$ along the diagonal of the BZ for $\lambda_v/t = 5$ ($\mathbb{Z}_2$-even phase) and $\lambda_v/t = 1$ ($\mathbb{Z}_2$-odd phase). (a) Trial functions are $|B;\uparrow_z\rangle$ and $|B;\downarrow_z\rangle$. (b) Trial functions are $|A;\uparrow_x\rangle$ and $|B;\downarrow_x\rangle$. . . 105

6.3. Wannier spreads Ω_I and $\tilde{\Omega}$ for the Kane-Mele model on a 60×60 $\mathbf{k}$-mesh, initialized using the trial functions of Eq. (6.12). “Initial” and “final” values are those computed before and after the iterative minimization respectively. The system is in the $\mathbb{Z}_2$-odd phase for $\lambda_v/t \lesssim 2.93$. 109

6.4. Hybrid Wannier centers $\bar{y}_n(k_x)$, in units of $\xi/2$, for the Kane-Mele model. $\mathbb{Z}_2$-even phase ($\lambda_v/t = 3$): (a) maxloc gauge; (b) WF gauge of Eq. (6.5). $\mathbb{Z}_2$-odd phase ($\lambda_v/t = 1$): (c) maxloc gauge; (d) WF gauge of Eq. (6.12). In each case, several periodic images are shown.114

7.1. Brillouin zone in 2D represented as (a) a torus, and (b) a cylinder. We choose the gauge discontinuity to be distributed along the cross-sectional cut of the torus that maps onto the end loops of the cylinder at $k_y = \pm\pi$. 118

7.2. (a) BZ in k-space. (b) States are parallel transported along $k_y = 0$, but are not periodic. (c) Periodicity is restored at $k_y = 0$. (d) Parallel transport of states at all k_x from $k_y = 0$ to $k_y = \pm\pi$. (d) Periodicity is restored at $k_x = 0$, and, hence, $k_x = 2\pi$, but not at other k_x. 126

7.3. Trajectory of V_{11} in the complex plane as k_x runs across the BZ, before (red dashed line) and after (solid black line) the global $\mathcal{U}_0$ rotation that minimizes $\mathcal{V}_{\mathrm{OD}}$ for a $\mathbb{Z}_2$-odd insulator. In neither case is the graph exactly a unit circle (dotted line), but $V_{11}(0) = 1$ and $V_{11}(\pi) = -1$. 134

Chapter 1
Introduction

Many problems in condensed matter physics allow for a reasonable solution within approximations that look quite drastic at the first sight. One of the most prominent examples in this respect is the success of the single-particle approximation in the treatment of many crystalline solids. It turns out that in a vast group of problems electron-electron and electron-nuclei interactions may be taken into account by means of some effective potential V_{eff} that captures the main effects of interactions in a mean-field manner. This allows one to reduce a complicated many-body problem to a problem of a single particle (quasi-electron) subject to the effective potential. There are different ways to construct V_{eff}, and the criterion of a successful construction is, of course, the agreement of the resultant single-particle description with experiment. A mean-field approach to crystalline solids holds the name of band theory [2–5].

In the case of a perfect crystalline system in the absence of external fields, the effective potential has the symmetry of the crystal lattice and thus is periodic. This makes it more convenient to work in the momentum representation assigning a reciprocal lattice vector (wave vector) $\mathbf{k}$ to the wavefunctions. Solution of a corresponding single-particle Schrödinger equation gives the allowed single-particle energy states as a function of wave vector. The Hamiltonian eigenvalues $\epsilon_n(\mathbf{k})$ form the so-called energy bands and a collection of them is called a band structure. Due to periodicity of the system, the quasi-electronic states with values of $\mathbf{k}$ within one unit cell of reciprocal space (Brillouin zone) give a complete description of the system.

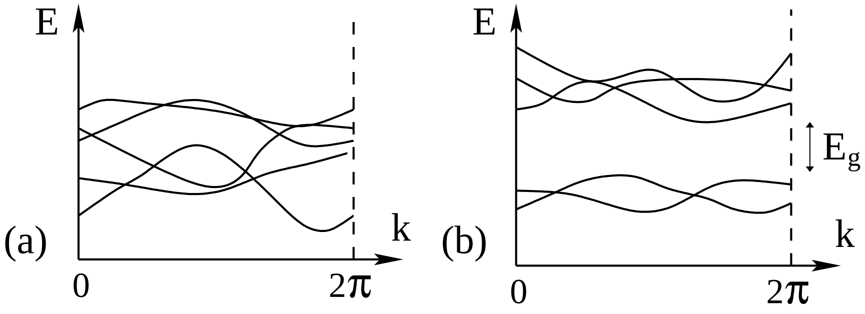

Figure 1.1: Possible band structures for a 1D system with two electrons per cell (lattice constant is taken to be a unit of length). Five lowest energy bands are plotted. (a) Metallic system. No energy gap. (b) Insulating system. The energy gap separates two occupied bands.

A huge amount of information about the system in question can be obtained from the band structure alone. For example, consider a 1D periodic system with two electrons per unit cell. Neglecting spin, one needs two bands to accommodate two spinless electrons. Possible band structures are illustrated in Fig. 1.1. In the band structure shown in panel (a) there are no two bands that can be completely separated from the rest. The two electrons can move freely from one band to another. Hence, we conclude that this system is metallic – an infinitesimal amount of energy is enough to drive the system away from its ground state. Quite the contrary, in the band structure of panel (b) the two lower bands (valence bands) are clearly separated from the the rest (conduction bands) by the forbidden energy region – an energy gap between the bottom of the conduction band and the top of the valence band. In the ground state the two lowest bands are occupied and the system is insulating, meaning that a finite amount (E_g) of energy is needed to excite an electron. Thus, we see that a mere glimpse at the band structure allows one to make predictions about the conducting properties of a given material and its possible response to external perturbations.

However, energy bands are not the only useful output of band theory. The geometric phases of the single-particle wavefunctions also turned out to be a very useful tool in band theory [6–9]. To name just a few applications, this phase is at the heart of the modern theory of polarization [10, 11] and anomalous contributions to the quantum Hall effect [12–14]. More recently, it was realized that the occupied states of a band insulator (see Fig. (1.1) (b)) can be viewed from a topological perspective [15–17]. Crystal periodicity allows one to consider the Brillouin zone as a torus, while the occupied electron wavefunctions form a Hilbert space at each point of this torus. Such structures are known to topology and go under the name of fibre bundles [18]. They may be classified by different topological invariants, thus leading to the conclusion that insulators might be topologically distinct. It turns out to be indeed the case – some insulators exhibit different properties from the others [1, 19], and this distinction cannot be guessed from the band structure, but only from the topology of the occupied wavefunctions [20–22]. For this reason these materials are denoted as “topological insulators” (TIs).

Two types of TIs have been discovered. First, it was pointed out that a two-dimensional (2D) insulator is characterized in general by a topological integer known as the “Chern number” or “TKNN index” [12]. A prospective insulator having a non-zero value of this integer would be known as a “Chern” or “quantum anomalous Hall” insulator. Such a material, when subject to an in-plane electric field, would exhibit quantization of the Hall conductivity in units of e^2/h (integer quantum Hall effect) even in the absence of a macroscopic transverse magnetic field, usually present in Hall-type experiments. The value of the Hall conductivity is related to the Chern number C via $\sigma_{xy} = Ce^2/h$. This phenomenon can be attributed to the existence of chiral edge states [23]. These states are usually understood as 1D wires that carry current along some preferred direction on the sample edge.

An explicit simple lattice model realizing a Chern insulator was devised [19], suggesting that real materials with such properties should exist. Since the Hall conductance is odd under the time-reversal ($\mathcal{T}$) operator, candidate materials would have broken $\mathcal{T}$ symmetry, e.g., they would be insulating ferromagnets. Despite the fact that these possibilities have been appreciated now for more than two decades, no known experimental realizations of a Chern insulator are yet known.

Second, a great deal of interest has surrounded the recent discovery of a different class of TIs known as $\mathbb{Z}_2$ insulators that realize the quantum spin Hall (QSH) effect [1, 24]. This effect is similar to the integer quantum Hall (IQH) effect discussed above, but the states at the edge are now helical. In the simplest case a helical edge state is composed of two counterpropagating chiral edge states that carry electrons of opposite spin. This results in no net charge transfer along the 1D wire, but there is an exactly quantized spin current instead. In a more general situation, spin-orbit coupling mixes different spin species, and spin current is not quantized any more. However, the absence or presence of spin-carrying channels at the edge of the sample is topologically protected.

Subsequent theoretical [25] and experimental [26] work has succeeded in identifying materials that realize the case of a QSH TI. Unlike the Chern index, which vanishes unless $\mathcal{T}$ is broken, the $\mathbb{Z}_2$ index (which takes values of 0 and 1, or equivalently, "even" and "odd") is only well defined when $\mathcal{T}$ is conserved. $\mathbb{Z}_2$ insulators are thus non-magnetic, although a spin-orbit or similar interaction is needed to mix the spins in a non-trivial way. Because $\mathcal{T}$ is preserved, the occupied states at $\mathbf{k}$ and $-\mathbf{k}$ form Kramers pairs (i.e., are mapped onto each other by $\mathcal{T}$), and one can associate a $\mathbb{Z}_2$ invariant with the way in which these Kramers pairs are connected across the Brillouin zone [27]. Being a topological invariant, the $\mathbb{Z}_2$ index cannot change along an adiabatic path that is everywhere gapped and $\mathcal{T}$-symmetric, and for this reason a $\mathbb{Z}_2$-even (normal) insulator cannot be connected to a $\mathbb{Z}_2$-odd (topological) one by such a path. In 2D there is a single $\mathbb{Z}_2$ invariant, and $\mathcal{T}$-invariant insulators are classified as "even" or "odd." TIs also exist in higher dimensions. A 3D $\mathcal{T}$-symmetric insulator is characterized by four $\mathbb{Z}_2$ invariants [28–30] and the classification is more complicated than in 2D. The non-trivial topology in this case results in metallic surfaces of the otherwise insulating sample.

In view of all this, there is an obvious motivation to develop simple yet effective methods for computing the topological indices of a given material. For the case of a Chern insulator the computation of the corresponding index is straightforward and consists of integration of some well defined quantity in the Brilloin zone. The case of $\mathbb{Z}_2$ insulators is much more complex. For centrosymmetric crystals, a convenient method was introduced in Ref. [31], where it was shown that the knowledge of the parity eigenvalues of the electronic states at only four $\mathcal{T}$-invariant momenta in 2D (or eight of them in 3D) is sufficient to compute the topological characteristics of a given material. This approach is limited to centrosymmetric systems, however, and the calculation of the $\mathbb{Z}_2$ invariant for noncentrosymmetric insulators is not so trivial.

The first question we address in this work is that of computing topological invariants in the absence of inversion symmetry. We develop a method that allows one to numerically obtain the $\mathbb{Z}_2$ invariant of a $\mathcal{T}$-symmetric insulator given the wavefunctions of the occupied states, which are the usual output of a band structure calculation. We illustrate the implementation of this method to real materials using density functional theory – the most successful scheme for constructing effective single particle Hamiltonians in band theory.

Another question that we consider in this work is related to a technique widely used in band theory. It consists of using Wannier functions (WFs) – functions that are exponentially localized in real space – instead of the Hamiltonian eigenstates for the description of various physical phenomena. This technique is extremely convenient for purposes like decomposing charge distributions, characterizing chemical bonding, computing electric polarization, calculating transport properties and constructing model Hamiltonians [7, 32–35].

In insulators the charge is localized, so it is natural to expect that the construction of an exponentially localized real-space representation should be straightforward. Starting with some particular choice of Bloch-like states $\{\tilde{\psi}_{n\mathbf{k}}\}$ that are not necessarily Hamiltonian eigenstates but span the occupied Hilbert space of an insulator, one can Fourier transform those states into the real-space representation. However, the localization properties of such states would strongly depend on the actual choice of the $\{\tilde{\psi}_{n\mathbf{k}}\}$, referred to as a gauge choice. In order to result in exponentially localized WFs, the gauge should be smooth in $\mathbf{k}$. For an ordinary insulator, it has been shown that such a choice is always possible [36].

For TIs, however, the situation is quite different. For example, in Chern insulators a non-zero topological invariant becomes an obstruction for choosing a smooth gauge. This, in turn, means that it is impossible to construct a set of exponentially localized WFs that would span the occupied space of a Chern insulator [37, 38]. But what about $\mathbb{Z}_2$ insulators? What is the effect of topology on the Wannier representation in this case? It turns out that a Wannier representation does exist in this case, but in order to construct it, $\mathcal{T}$ symmetry should be broken in the gauge in some specific manner. In other words, if in the momentum representation one chooses the wavefunctions representing the occupied space at $\mathbf{k}$ and $-k$ to be $\mathcal{T}$-images of one another, then the Wannier construction necessarily breaks down. This is very different from what happens in the ordinary $\mathcal{T}$-symmetric insulators, where the WFs respect the symmetry of the system and come in Kramers pairs. Thus, in the case of $\mathbb{Z}_2$ insulators, any smooth gauge violates the inherent $\mathcal{T}$ symmetry of the system.

The fact that the gauge violates the inherent $\mathcal{T}$ symmetry of the system might seem confusing, but there is no contradiction here. The reason is that different gauge choices correspond to the same physics. Violation of $\mathcal{T}$ symmetry in the gauge does *not* imply violation of $\mathcal{T}$ symmetry in the Hamiltonian. The situation is similar to the gauge choice for a vector potential $\mathbf{A}$ in electromagnetism. For example, for an atom in a uniform magnetic field along $\hat{\mathbf{z}}$, one is perfectly free to choose $\mathbf{A} = (B_0/2)(-y, x)$ or $\mathbf{A} = B_0(-y, 0)$; only the former has rotational symmetry in the gauge, but the physical predictions must be identical whichever one is used. Thus, a breaking of symmetry in the gauge, while it may complicate some calculations or intermediate results, cannot ultimately appear in the form of any physical symmetry breaking.

In the present work we describe an explicit construction of WFs for a model $\mathbb{Z}_2$ TI, breaking $\mathcal{T}$ symmetry in the gauge in a very particular manner. We then set up a general theory that explains how a smooth gauge can be constructed for $\mathbb{Z}_2$ TIs and puts the necessity of $\mathcal{T}$-breaking into a rigorous framework. These results allow for the generalization of the WF technique to topologically non-trivial band structures.

This thesis is organized as follows. Chapters 2, 3 and 4 introduce the background material, setting the stage for the material presented in the rest of the work. Chapter 2 reviews basic results of the band theory of solids and introduces two methods of doing electronic structure calculations: the tight-binding approximation and density functional theory. In Chapter 3 TIs are introduced is some detail, including models with which we illustrate further material. An introduction to WFs is given in Chapter 4. The remainder of the thesis represents original research. Chapter 5 describes a method for computing the $\mathbb{Z}_2$ topological invariant for non-centrosymmetric systems and contains a discussion of how to implement this method in the density functional framework, along with some examples. This work is based on Ref. [39]. Chapters 6 and 7 are dedicated to the question of the Wannier representation construction for $\mathbb{Z}_2$ insulators. First, an explicit construction of WFs for a particular model TI is presented in Chapter 6 (based on Ref. [40]), and then the method is put on a general footing in Chapter 7 (Ref. [41]). The conclusions and outlook are presented in the last chapter of the thesis.

Chapter 2

Basic notions of band theory

To set the stage for the following discussion, we present a quick review of some results and methods of band theory. In particular, we discuss the reduction of the original many-body problem to a set of single-particle equations, concentrating on a particular method of obtaining these equations, namely density functional theory (DFT). In many cases DFT is sufficient to get good agreement with experiment, thus being a powerful numerical tool for calculating realistic band structures and related quantities. On the other extreme of theoretical approaches to solids is the tight-binding approximation (TBA), which provides a good qualitative understanding of the physics involved within a simplified model. We discuss some details of the TBA in the last section of the present chapter.

2.1 From the many-body problem to many single-body problems

The whole area of condensed matter physics in all its diversity and complexity is rooted to a relatively simple Hamiltonian

$$\hat{H} = -\sum_{j}^{N_{\mathrm{e}}} \frac{\hbar^2}{2m_{\mathrm{e}}} \nabla^2_{\mathbf{r}_j} - \sum_{\ell}^{N_{\mathrm{n}}} \frac{\hbar^2}{2M_\ell} \nabla^2_{\mathbf{R}_\ell} + V_{\mathrm{ee}} + V_{\mathrm{nn}} + V_{\mathrm{en}} + V_{\mathrm{SO}}, \tag{2.1}$$

which describes a system of N_{e} electrons and N_{n} nuclei. Here $\mathbf{R}_\ell$ and M_ℓ stand for nuclear coordinates and masses, while $\mathbf{r}_j$ refers to electrons with mass m_{e}. The first two terms of Eq. (2.1) correspond to the kinetic energy of electrons and ions respectively, and the next three stand for electron-electron, nucleus-nucleus and electron-nucleus Coulomb interaction, in accord with the labeling. The last term refers to the spin-orbit interaction. Other terms, such as couplings to external fields, might also appear, but they usually can be treated perturbatively and are omitted for the purposes of the present overview.

In case of a real solid this Hamiltonian describes a many-body system of a large number ($> 10^{23}$ for a crystal) of particles and, in general, does not allow for an exact solution. It is mainly the electron-electron interaction term that makes the problem so diverse and complex, being a truly many-body term. The other two Coulomb terms also represent many-body interactions, but the complications they cause can be avoided by a set of approximations that has proven to work very well for a large class of problems. Let us consider these approximations in detail.

(A) *Frozen nuclei approximation:* In this approximation all the nuclear coordinates are considered to be fixed in their equilibrium positions and the dynamics of nuclei is completely disregarded. Although for some problems this limitation is too strict, it works very well for most cases.

(A') *Adiabatic approximation:* Now the nuclei are allowed to oscillate around their equilibrium positions, thus giving rise to phonons and other lattice-mediated phenomena. However, the nuclear motion is considered to be slow compared to the that of the electrons, and the electronic wavefunction is treated as a function of the instantaneous nuclear positions. That is, the electrons are assumed to follow the nuclei instantaneously, remaining in the same state of the electronic Hamiltonian with the nuclear coordinates treated as parameters. This approximation allows one to decouple the electronic and nuclear degrees of freedom and solve for $\Psi(\{\mathbf{r}_j\}, \{\mathbf{R}_\ell\}, t) = \chi(\{\mathbf{R}_\ell\}, t)\Phi_R(\{\mathbf{r}_j\})$, with $\Phi_R(\{\mathbf{r}_j\})$ being the ground-state electronic wavefunction for the current nuclear configuration [4, 42].

By means of these two approximations it becomes possible to get rid of the complications due to the many-body nature of V_{nn} and V_{en}. The nucleus-nucleus interaction can be reduced to a constant term closely related to the Madelung energy [2]. The electron-nucleus contribution becomes a one-body operator, i.e., describes one electron moving in the field of immobile positive charges.

It turns out that in many cases V_{ee} can also be reduced to a one-body form by means of the following approximation.

(B) *Mean field approximation:* Some effective field is introduced in order to mimic the many-body effects of the electron-electron operator, and V_{ee} is replaced by V_{eff}, which is a single-particle operator. This is an extremely powerful approximation that serves as a basis for many successful effective theories. The resultant single-particle Hamiltonian takes the form

$$H_{\mathrm{sp}} = -\frac{\hbar^2}{2m}\nabla^2_{\mathbf{r}} + V_{\mathrm{eff}}(\mathbf{r}) + V_{\mathrm{SO}}, \tag{2.2}$$

where the interaction of the electron with the lattice V_{en} is now absorbed in V_{eff} and the constant Madelung term is dropped. Note that what is now referred to as an “electron” is not really the original electron of Eq. (2.1), but rather a “quasi-particle” that moves in the mean effective field of electrons and nuclei.

Now the initial many-body problem is replaced with a set of single-particle equations of the form

$$H_{\mathrm{sp}}|\phi_i\rangle = \epsilon_i|\phi_i\rangle. \tag{2.3}$$

for each quasi-particle. For the moment, let us neglect spin-orbit coupling and discuss some properties of the single-particle Hamiltonian obtained above. Since quasi-particles are non-interacting, the many-body wavefunction of N such particles $\langle\mathbf{r}|\Psi\rangle = \Psi(\mathbf{r}_1, .., \mathbf{r}_n)$ should have the form of a product of single-particle eigenstates $|\phi_i\rangle$. However, since quasi-particles have the fermionic nature inherent to the original electrons, the wavefunction should be antisymmetric with respect to exchange of coordinates of any two particles. This is achieved by means of the antisymmetrizing operator $\hat{A}$ which allows the many-body wavefunction to be written in the convenient form of a Slater determinant:

$$\Psi(\mathbf{r}_1, .., \mathbf{r}_N) = \hat{A} \prod_{j=1}^{N} \phi_j(\mathbf{r}_j) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\mathbf{x}_1) & .. & \phi_N(\mathbf{x}_1) \\ : & : & : \\ \phi_1(\mathbf{x}_N) & .. & \phi_N(\mathbf{x}_N) \end{vmatrix} \tag{2.4}$$

where $\mathbf{x}$ stands for both coordinate $\mathbf{r}$ and spin σ.

When spin-orbit coupling is taken into account the single-particle wavefunction becomes a two-component spinor

$$\psi_j(\mathbf{r}) = \begin{pmatrix} \phi_{j\uparrow}(\mathbf{r}) \\ \phi_{j\downarrow}(\mathbf{r}) \end{pmatrix}, \tag{2.5}$$

where each of the spinor components is a complex function. One can see that a spinor can describe noncollinear spins, e.g., the spinor $(1, 0)^T$ corresponds to a state with a spin direction along the positive $\hat{\mathbf{z}}$-axis, while $(1, 1)^T/\sqrt{2}$ corresponds to a state with a spin in the $+\hat{\mathbf{x}}$ direction. Note the normalization factor in the latter case – we consider spinors to be normalized to unity.

So far, our considerations have been quite general, without any specific reference to the periodicity of atomic arrangement in crystalline solids. However, lattice periodicity of the ionic (nuclear) potential results in lattice periodicity of the single-particle effective potential. This turns out to be a great simplification of the problem, since, as we show below, it is sufficient to solve the Schrödinger equation (2.3) only in a limited portion of space.

2.1.1 Crystal potential and Bloch theorem

A perfect crystal may be represented by a building block – a unit cell, which is periodically repeated over a lattice in space. In m dimensions the lattice is formed by m lattice vectors $\mathbf{a}_i$, and due to periodicity of the structure, a general vector that connects one cell to its periodic copy is, in 3D,

$$\mathbf{R}_n = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3, \tag{2.6}$$

where $n = n_1, n_2, n_3$.

As was mentioned above, the potential $V_{\text{eff}}(\mathbf{r})$ should retain lattice periodicity. When dealing with periodic functions, the Fourier transform becomes invaluable; it maps a periodic function of $\mathbf{r}$ into a reciprocal lattice, which is a counterpart of the crystalline lattice in momentum space. Reciprocal lattice basis vectors are defined by the equation

$$\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}. \tag{2.7}$$

It is possible to write down the explicit form of $\mathbf{b}_j$ in term of the real-space basis vectors $\mathbf{a}_i$. In 1D, for a crystal with a lattice constant a this form is obvious:

$$b = 2\pi/a. \tag{2.8}$$

In 2D the area of the unit cell is introduced $A = \hat{\mathbf{n}} \times \mathbf{a}_1 \cdot \mathbf{a}_2 = a_1 \wedge a_2$, where $\hat{\mathbf{n}}$ is a unit vector normal to the surface of the crystal, and the reciprocal lattice vectors become

$$\begin{aligned} \mathbf{b}_1 &= \frac{2\pi}{A}\mathbf{a}_2 \times \hat{\mathbf{n}}, \\ \mathbf{b}_2 &= \frac{2\pi}{A}\hat{\mathbf{n}} \times \mathbf{a}_1. \end{aligned} \tag{2.9}$$

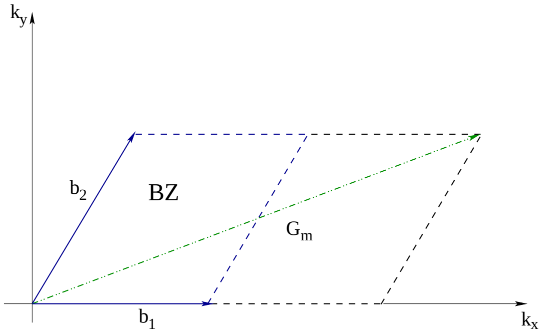

Figure 2.1: Reciprocal space (**k**-space). Brillouin zone (shown in blue) is formed by the primitive reciprocal lattice vectors $\mathbf{b}_1$ and $\mathbf{b}_2$. A non-primitive reciprocal lattice vector $\mathbf{G}_m$ is shown in green.

In 3D the volume is $V = \mathbf{a}_1 \cdot \mathbf{a}_2 \times \mathbf{a}_3 = a_1 \wedge a_2 \wedge a_3$ and

$$\begin{aligned} \mathbf{b}_1 &= \frac{2\pi}{V}\mathbf{a}_2 \times \mathbf{a}_3, \\ \mathbf{b}_2 &= \frac{2\pi}{V}\mathbf{a}_3 \times \mathbf{a}_1, \\ \mathbf{b}_3 &= \frac{2\pi}{V}\mathbf{a}_1 \times \mathbf{a}_2. \end{aligned} \tag{2.10}$$

The whole reciprocal lattice is formed by reciprocal lattice vectors of the form

$$\mathbf{G}_m = m_1\mathbf{b}_1 + m_2\mathbf{b}_2 + m_3\mathbf{b}_3 \tag{2.11}$$

just as the real-space lattice was formed by vectors (2.6).

Figure 2.1 illustrates the reciprocal lattice in 2D. A unit cell in reciprocal space is called a Brillouin zone (BZ), and as in the case of a direct lattice is not uniquely defined. In what follows, we choose the BZ to be a unit cell of minimum possible volume, as shown in Fig. 2.1. Finally, note that for any pair $\mathbf{R}_n$ and $\mathbf{G}_m$ the relation

$$\mathbf{R}_n \cdot \mathbf{G}_m = 2\pi\ell \tag{2.12}$$

holds, where $\ell = n_1 m_1 + n_2 m_2 + n_3 m_3$ is an integer, so that

$$e^{i\mathbf{G}_m \cdot \mathbf{R}_n} = 1. \tag{2.13}$$

Any lattice-periodic function $f(\mathbf{r})$ can thus be Fourier transformed in terms of reciprocal lattice vectors as

$$f(\mathbf{r}) = \sum_{\mathbf{G}} e^{i\mathbf{G}\cdot\mathbf{r}} f(\mathbf{G}) \tag{2.14}$$

This is a useful result, since as we will show now, the eigenstates of the single-particle Hamiltonian can be written in terms of lattice-periodic functions, so that the whole machinery of Fourier analysis is conveniently invoked in band theory.

Bloch theorem

We will now prove the basic theorem of band theory. It says that when the effective potential is periodic, $V_{\text{eff}}(\mathbf{r}) = V_{\text{eff}}(\mathbf{r} + \mathbf{R})$, the eigenstates of the Hamiltonian (2.2) are periodic up to a phase factor:

$$\psi_{n\mathbf{k}}(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}} \psi_{n\mathbf{k}}(\mathbf{r}). \tag{2.15}$$

For now, $\mathbf{k}$ is just an extra label on the wavefunction. We will clarify its meaning below. Equation (2.17) can be written differently as

$$\psi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n\mathbf{k}}(\mathbf{r}), \tag{2.16}$$

where

$$u_{n\mathbf{k}}(\mathbf{r} + \mathbf{R}) = u_{n\mathbf{k}}(\mathbf{r}). \tag{2.17}$$

The states $\psi_{n\mathbf{k}}$ ($u_{n\mathbf{k}}$) are called Bloch states (cell-periodic Bloch states).

To prove this theorem, let us consider the translation operator $\hat{T}_{\mathbf{R}_n}$, which acts on functions of $\mathbf{r}$ according to the rule

$$\hat{T}_{\mathbf{R}_n} f(\mathbf{r}) = f(\mathbf{r} - \mathbf{R}_n). \tag{2.18}$$

Different translation operators commute with each other, $[\hat{T}_{\mathbf{R}_n}, \hat{T}_{\mathbf{R}_m}] = 0$, and form an Abelian group: for each $\hat{T}_{\mathbf{R}_n}$ the unit element is given by $\hat{T}_{\mathbf{0}}$ and the inverse is $\hat{T}_{-\mathbf{R}_n}$. The group operation for two elements of the group obviously gives another element of the group:

$$\hat{T}_{\mathbf{R}_n} \hat{T}_{\mathbf{R}_m} = \hat{T}_{\mathbf{R}_n + \mathbf{R}_m}. \tag{2.19}$$

Since the potential $V_{\text{eff}}(\mathbf{r})$ has the periodicity of the lattice, the translation operators commute with the single-particle Hamiltonian, i.e., $[\hat{T}_{\mathbf{R}_n}, H_{\text{sp}}] = 0$. This allows one to choose the Hamiltonian eigenstates to be simultaneously the eigenstates of $\hat{T}_{\mathbf{R}_n}$. We have

$$H_{\text{sp}} |\psi_{n\mathbf{k}}\rangle = \epsilon_n(\mathbf{k}) |\psi_{n\mathbf{k}}\rangle \tag{2.20}$$

and

$$\hat{T}_{\mathbf{R}_m} |\psi_{n\mathbf{k}}\rangle = c_{\mathbf{R}_m} |\psi_{n\mathbf{k}}\rangle, \tag{2.21}$$

where $c_{\mathbf{R}_m}$ is a complex eigenvalue of $\hat{T}_{\mathbf{R}_m}$. In accord with Eq. (2.19) for the product of $\hat{T}_{\mathbf{R}_m}$ and $\hat{T}_{\mathbf{R}_h}$, the relation

$$\hat{T}_{\mathbf{R}_m} \hat{T}_{\mathbf{R}_h} |\psi_{n\mathbf{k}}\rangle = c_{\mathbf{R}_m} c_{\mathbf{R}_h} |\psi_{n\mathbf{k}}\rangle = c_{\mathbf{R}_m + \mathbf{R}_h} |\psi_{n\mathbf{k}}\rangle \tag{2.22}$$

is valid. The only complex function of $\mathbf{R}$ that can satisfy such an equation is an exponential. We thus set $c_{\mathbf{R}} = e^{-i\mathbf{k}\cdot\mathbf{R}}$, so that

$$\hat{T}_{\mathbf{R}_n} \psi_{n\mathbf{k}}(\mathbf{r}) = \psi_{n\mathbf{k}}(\mathbf{r} - \mathbf{R}_n) = e^{-i\mathbf{k}\cdot\mathbf{R}_n} \psi_{n\mathbf{k}}(\mathbf{r}). \tag{2.23}$$

It is now straightforward to write down the eigenfunctions of $\hat{T}_{\mathbf{R}_n}$. Using Eq. (2.12), one can check that $e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}$ satisfies the eigenvalue equation for the translation operator. Indeed,

$$\hat{T}_{\mathbf{R}_n} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} = e^{i(\mathbf{k}+\mathbf{G})\cdot(\mathbf{r}-\mathbf{R}_n)} = e^{-i\mathbf{k}\cdot\mathbf{R}_n} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}} = c_{\mathbf{R}_n} e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}. \tag{2.24}$$

This means that we can expand an eigenstate of the single-particle Hamiltonian (2.2) in terms of eigenfunctions of $\hat{T}_{\mathbf{R}_n}$ (plane waves) that correspond to the same eigenvalue. That is,

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} C_{n\mathbf{k}}(\mathbf{G}) e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{r}}, \tag{2.25}$$

which proves the Bloch theorem. To see that, we just rewrite the above equation as $\psi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n\mathbf{k}}(\mathbf{r})$, where

$$u_{n\mathbf{k}}(\mathbf{r}) = \sum_{\mathbf{G}} C_{n\mathbf{k}}(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}}, \tag{2.26}$$

and, obviously, $u_{n\mathbf{k}}(\mathbf{r}+\mathbf{R}) = u_{n\mathbf{k}}(\mathbf{r})$.

2.1.2 Reciprocal space, Bloch states, and band structure

Let us now discuss the physical meaning of the label $\mathbf{k}$. As one might have noticed already, $\mathbf{k}$ plays the same role in reciprocal space as $\mathbf{r}$ does in direct space. First, notice that since the exponential $e^{i\mathbf{k}\cdot\mathbf{R}}$ has to be dimensionless, $\mathbf{k}$ has dimensions of inverse length, just as reciprocal lattice vectors $\mathbf{G}$ do. Hence, we can consider an eigenstate of the Hamiltonian with $\mathbf{k}$ translated by $\mathbf{G}$. This eigenstate

$$\psi_{n,\mathbf{k}+\mathbf{G}}(\mathbf{r}+\mathbf{R}) = \psi_{n,\mathbf{k}+\mathbf{G}}(\mathbf{r}) e^{i(\mathbf{k}+\mathbf{G})\cdot\mathbf{R}} = e^{i\mathbf{k}\cdot\mathbf{R}} \psi_{n,\mathbf{k}+\mathbf{G}}(\mathbf{r}) \tag{2.27}$$

obeys exactly the same boundary condition as $\psi_{n\mathbf{k}}$ does, suggesting that $|\psi_{n\mathbf{k}}\rangle$ and $|\psi_{n,\mathbf{k}+\mathbf{G}}\rangle$ are duplicate labels for the same state. Hence, it is justified to

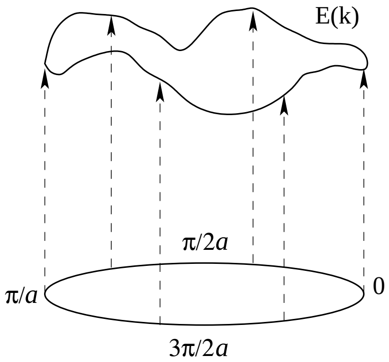

Figure 2.2: Energy band in 1D k-space.

consider the behavior of $\psi_{n\mathbf{k}}$ only in the BZ (say, $k \in [0, 2\pi/a]$ or $k \in [-\pi/a, \pi/a]$ in 1D), so that $\mathbf{k} = \sum_j \kappa_j \mathbf{b}_j$, where $\mathbf{b}_j$ are primitive reciprocal lattice vectors.

The fact that we are interested in **k** only within one BZ means that **k**-space may be regarded as a closed manifold. Indeed, the BZ is a unit cell in reciprocal space, and its boundaries are periodic images of each other, so that they can be identified. For example, a circle S^1 represents **k**-space in case of 1D, and instead of considering periodic functions of **k** in the whole reciprocal space, we might equivalently consider functions defined on a circle.

As an example, consider the energy levels of the singe-particle Schrödinger equation (2.20), called energy bands. They are functions of **k**, and $\epsilon_n(\mathbf{k}+\mathbf{G}) = \epsilon_n(\mathbf{k})$, so that a collection of bands has a **G**-periodic dependence on **k**, and a set of bands with **k** within the BZ is referred to as band structure. Treating the BZ as a closed manifold allows one to define energy bands as a mapping from this manifold into real numbers. For instance, in 1D a band n represents a mapping $\epsilon_{n\mathbf{k}} : S^1 \to \mathbb{R}^1$, as illustrated in Fig. 2.2. In 2D $\mathbf{k} = (k_x, k_y)$ and the BZ is represented by a 2-torus $T^2 = S^1 \times S^1$ (see Fig. 2.3), and the band $\epsilon_{n\mathbf{k}} : T^2 \to \mathbb{R}^1$. Continuing along these lines, a 3D BZ is represented by a 3-torus $T^3 = S^1 \times S^1 \times S^1$.

As we have seen above, the single-particle wavefunctions at **k** and $\mathbf{k} + \mathbf{G}$

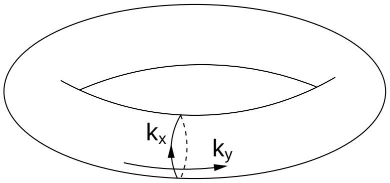

Figure 2.3: BZ in 2D is a torus.

correspond to the same physical state, and thus can differ only by a phase factor. Taking on the view of a BZ as a closed manifold, we fix this arbitrary phase factor to be unity, so that

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \langle \mathbf{r}|\psi_{n\mathbf{k}}\rangle = \psi_{n,\mathbf{k}+\mathbf{G}}(\mathbf{r}). \tag{2.28}$$

In terms of cell-periodic Bloch functions $u_{n\mathbf{k}}$ this condition reads

$$|u_{n,\mathbf{k}+\mathbf{G}}\rangle = e^{-i\mathbf{G}\cdot\mathbf{r}}|u_{n\mathbf{k}}\rangle. \tag{2.29}$$

Thus, the u-functions are periodic in real space, but not in reciprocal space, while for ψ-functions the reverse is true.

In what follows we deal with $u_{n\mathbf{k}}$ rather than $\psi_{n\mathbf{k}}$, since they are better behaved. In particular, the derivatives $\frac{d}{dk}|u_{n\mathbf{k}}\rangle$ are well-behaved and belong to the same Hilbert space as $|u_{n\mathbf{k}}\rangle$. This is not the case for $|\psi_{n\mathbf{k}}\rangle$, since

$$\frac{d}{dk}\psi_{nk}(r) = \frac{d}{dk}\left(e^{ikr}u_{nk}(r)\right) = e^{ikr}\frac{d}{dk}u_{nk}(r) + ire^{ikr}u_{nk}(r)$$

and the second term obviously blows up at large r, since for $\psi_{nk}(r)$ the real-space argument spans the whole space.

There is one seeming drawback of the u-functions, namely that unlike the ψ-functions they are not eigenfunctions of the Hamiltonian. However, this issue is easily solved by transforming H_{sp} to

$$H_{\mathbf{k}} = e^{-i\mathbf{k}\cdot\mathbf{r}} H_{\text{sp}} e^{i\mathbf{k}\cdot\mathbf{r}}. \tag{2.30}$$

This leads to the effective Schrödinger equation

$$H_{\mathbf{k}} u_{n\mathbf{k}}(\mathbf{r}) = \left[\frac{1}{2m}\left(\mathbf{p} + \hbar\mathbf{k}\right)^2 + V_{\text{eff}}(\mathbf{r})\right] u_{n\mathbf{k}}(\mathbf{r}) = \epsilon_n(\mathbf{k}) u_{n\mathbf{k}}(\mathbf{r}). \tag{2.31}$$

When the spin-orbit interaction

$$V_{\text{SO}} = \frac{\hbar}{4m^2c^2}\mathbf{p}\cdot\boldsymbol{\sigma}\times\left(\boldsymbol{\nabla}V_{\text{eff}}(\mathbf{r})\right) \tag{2.32}$$

is included in the calculation, one can write a similar Hamiltonian for spinor wavefunctions [43]:

$$H_{\mathbf{k}} = \left[\frac{1}{2m}\left(\mathbf{p}^2 + \hbar^2\mathbf{k}^2\right) + \frac{\hbar}{m}\mathbf{k}\cdot\boldsymbol{\mathcal{P}} + V_{\text{eff}}(\mathbf{r}) + \frac{\hbar}{4m^2c^2}\mathbf{p}\cdot\boldsymbol{\sigma}\times\left(\boldsymbol{\nabla}V_{\text{eff}}(\mathbf{r})\right)\right], \tag{2.33}$$

where

$$\boldsymbol{\mathcal{P}} = \mathbf{p} + \frac{\hbar}{4mc^2}\boldsymbol{\sigma}\times\left(\boldsymbol{\nabla}V_{\text{eff}}\right). \tag{2.34}$$

Note that for the case of a spherically symmetric potential, the term $\boldsymbol{\sigma}\times(\boldsymbol{\nabla}V_{\text{eff}})$ is proportional to $\mathbf{L}\cdot\mathbf{S}$, where $\mathbf{L}$ and $\mathbf{S}$ stand for angular momentum and spin operators correspondingly.

The solution of the Schrödinger equation with this Hamiltonian provides both the band structure and the Bloch wavefunctions. Usually, in applications to crystalline solids one is given a number of electrons, whose behavior within the unit cell is described by renormalized quasi-particles. In what follows we will see that it is usually the valence electrons – the ones from the outer occupied atomic shells – that are of interest in solid state applications, since the electrons of the deeper filled shells (core electrons) remain almost unaltered in the presence of the

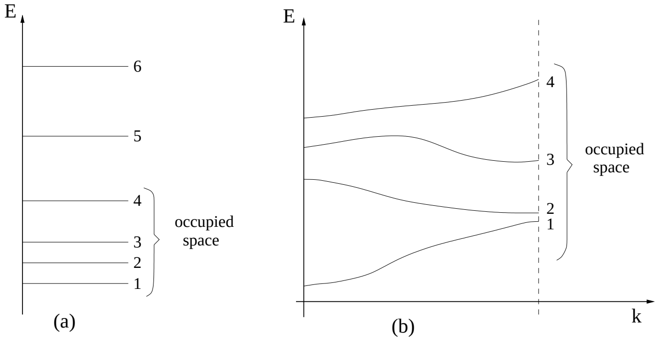

Figure 2.4: Energy levels of valence electrons (a) form a band structure in crystalline environment (b). Core states are lower in energy.

crystal potential and are tightly bound to the nucleus, forming an ion.

Figure 1.1 illustrates possible band structures of a 1D crystal. Assuming two electrons per unit cell, we see that the ground state of the system in panel (b) is represented by a set of two bands $\epsilon_n(\mathbf{k})$ for $n = 1, 2$, while the other bands are separated from this set by an energy gap. This separation of the manifold of occupied states is characteristic of insulators. In real applications, the bands usually shown in the band structure illustrations are those of the valence electrons. The atomic levels of valence electrons are modified by the potential of the crystalline environment and acquire dispersion, as shown in Fig. 2.4. The core states usually lie far below the valence states, and are thus of little interest, having almost no influence on the physical properties of a crystal.

Now that we have seen the potential of the single-particle approach, it is time to come back to the question of constructing V_{eff}. There are many possible approaches to the problem, and we proceed to discuss one of the most powerful of those known so far – DFT.

2.2 Density functional theory

The construction of the effective single-particle potential was based on some drastic assumptions about the many-body ground state of the system [2] until Hohenberg and Kohn [44] proved that any property of the many-electron system can be written as a functional of the single-particle electron density

$$n_0(\mathbf{r}) = N \int \Psi_0^*(\mathbf{r}, \mathbf{r}_2, .., \mathbf{r}_N)\Psi_0(\mathbf{r}, \mathbf{r}_2, .., \mathbf{r}_N) d\mathbf{r}_2..d\mathbf{r}_N \tag{2.35}$$

obtained from the many-body ground state Ψ_0, setting up what is now called DFT. In its original formulation DFT is an exact theory of many-electron systems, which allows one to write the total energy as a functional of the single-particle density $n(\mathbf{r})$, and the minimum of this functional occurs when $n(\mathbf{r}) = n_0(\mathbf{r})$. However, to date, the exact form of this functional is unknown, and in this respect, DFT presents merely an alternative view at the original problem (2.1). What makes this approach so useful is the possibility to construct a good approximate mean-field solution using a standard scheme, suggested by Kohn and Sham [45]. We will first review the basic equations of a general DFT formalism and then proceed to the Kohn-Sham (KS) iterative solution scheme used nowadays in numerical codes.

2.2.1 General formalism of DFT

For a moment, let us neglect the spin-orbit interaction and consider an electronic Hamiltonian

$$H_{\mathrm{el}} = -\frac{\hbar^2}{2m_{\mathrm{e}}} \sum_j \boldsymbol{\nabla}_j + \frac{1}{2} \sum_{i \neq j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + V_{\mathrm{ext}} \tag{2.36}$$

with the external potential of the form

$$V_{\mathrm{ext}} = \int v_{\mathrm{ext}}(\mathbf{r}) n(\mathbf{r}) d\mathbf{r}, \tag{2.37}$$

where the potential density $v_{\text{ext}}(\mathbf{r})$ is introduced. The nuclear potential V_{en} is a particular example of such an external potential. The whole theory resides on two theorems due to Hohenberg and Kohn[44] (for the proof see, for example, Ref. [5]):

(1) The external potential $v_{\text{ext}}(\mathbf{r})$ (up to a constant term) stands in one-to-one correspondence with the ground state electron density $n_0(\mathbf{r})$. Thus, the many-body wavefunction is a functional of electron density in principle, since it is determined by $v_{\text{ext}}(\mathbf{r})$.

(2) One can define a universal energy functional $E^{[v_{\text{ext}}]}[n]$ of the electron density $n(\mathbf{r})$,

$$E^{[v_{\text{ext}}]}[n] = F[n] + \int v_{\text{ext}}(\mathbf{r})n(\mathbf{r})d\mathbf{r}, \tag{2.38}$$

where $F[n] = \langle\Psi[n]|T + V_{\text{ee}}|\Psi[n]\rangle$, T being the kinetic energy. The ground state energy of the system of interacting electrons is then obtained by finding the global minimum of this energy functional for the given v_{ext}. The density that corresponds to the minimum is exactly the ground state density $n_0(\mathbf{r})$.

The theory presented so far is exact. It gives an alternative approach to the many-body problem of interacting electrons in an external field. However, as is, the problem does not look any simpler then the solution of the full Schrödinger equation with the Hamiltonian (2.1). It is the ansatz solution suggested by Kohn and Sham [45] that made a DFT approach feasible.

2.2.2 Kohn-Sham solution scheme

The particular solution of the problem of minimizing the energy functional (2.38) that we are going to discuss is due to Kohn and Sham [45]. The idea is to find a non-interacting system that has the same particle density as the interacting one. A many-body state for non-interacting fermions takes the form of a Slater determinant (2.4) of single-particle orbitals $\phi_j(\mathbf{r})$, where j stands for a collective index for all the quantum numbers. The density in this case is

$$n(\mathbf{r}) = \sum_j |\phi_j(\mathbf{r})|^2, \tag{2.39}$$

where the summation involves occupied orbitals only. The energy functional takes the form

$$E[n] = -\frac{\hbar^2}{2m_\mathrm{e}} \sum_j \langle\phi_j|\boldsymbol{\nabla}|\phi_j\rangle + \iint \frac{n(\mathbf{r})n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}d\mathbf{r}' + \int v_\mathrm{ext}(\mathbf{r})n(\mathbf{r})d\mathbf{r} + E^\mathrm{XC}[n], \tag{2.40}$$

where the last term is the exchange-correlation functional, which captures the effects of interactions. The exact form of this term is unknown, and different approximate expressions are usually used, as we discuss below. The reduction to single-particle equations of the form of Eq. (2.3) is done by varying the energy functional with respect to ϕ_j^*, so that

$$\delta n(\mathbf{r}) = \delta\phi_j^*(\mathbf{r})\phi_j(\mathbf{r}), \tag{2.41}$$

with a condition that the total number of particles is conserved [2, 45]

$$\int \delta n(\mathbf{r})d\mathbf{r} = 0. \tag{2.42}$$

Using Lagrange multipliers ϵ_j to account for this restriction, one arrives at a single-particle equation

$$\left[-\frac{\hbar^2}{2m_\mathrm{e}}\boldsymbol{\nabla} + V_\mathrm{eff}(\mathbf{r})\right]\phi_j = \epsilon_j\phi_j \tag{2.43}$$

with

$$V_\mathrm{eff}(\mathbf{r}) = \int \frac{n(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} d\mathbf{r}' + v_\mathrm{ext}(\mathbf{r}) + \mu^\mathrm{XC}, \tag{2.44}$$

where

$$\mu^{\mathrm{XC}} = \frac{\delta E^{\mathrm{XC}}[n(\mathbf{r})]}{\delta n(\mathbf{r})} \tag{2.45}$$

is the variational functional derivative of the exchange-correlation functional, which we will discuss in some details in the next paragraph. The set of single-particle equations (2.43) is called the Kohn-Sham equations, and the single-particle orbitals ϕ_j are called Kohn-Sham orbitals.

The effective potential in the KS equations is a function of the density, which in turn is constructed out of the single-particle states. These type of problems are solved iteratively. In this particular case the solution goes in the following steps.

(1) Make some initial guess for the effective potential.

(2) Solve the single particle equation (2.43) to find ϕ_j.

(3) Calculate the charge density $n(\mathbf{r}) = \sum_{\mathbf{j}} |\phi_{\mathbf{j}}(\mathbf{r})|^{\mathbf{2}}$.

(4) Construct new $V_{\mathrm{eff}}^{\mathrm{new}}$ from this density via Eqs. (2.44 - 2.45) and compare it to the previously used one.

(5) If $|V_{\mathrm{eff}}^{\mathrm{new}} - V_{\mathrm{eff}}^{\mathrm{old}}| > \alpha_{\mathrm{threshold}}$, go to step *(2)* using $V_{\mathrm{eff}} = V_{\mathrm{eff}}^{\mathrm{new}}$.

(6) Proceed until convergence is reached. For example, $|V_{\mathrm{eff}}^{\mathrm{new}} - V_{\mathrm{eff}}^{\mathrm{old}}| < \alpha_{\mathrm{threshold}}$ twice in a row.

Note that the effective potential (2.44) is exact once the exact form of E^{XC} is provided. In that case the above iteration scheme would in principle converge to the correct answer for any system of interacting electrons, in particular for any material. However, the problem of finding an exact form of E^{XC} is not solved, and one has to use different approximations, which we now discuss.

2.2.3 Exchange-correlation potential

In the simplest approach the exchange-correlation energy functional can be approximated by a local functional of density

$$E^{\mathrm{XC}}[n] = \int n(\mathbf{r})\epsilon^{\mathrm{XC}}(n(\mathbf{r}))d\mathbf{r} = \int n(\mathbf{r})\left(\epsilon^{\mathrm{X}}(n(\mathbf{r})) + \epsilon^{\mathrm{C}}(n(\mathbf{r}))\right)d\mathbf{r}, \qquad (2.46)$$

where $\epsilon^{\mathrm{XC}} = \epsilon^{\mathrm{X}} + \epsilon^{\mathrm{C}}$ is the exchange-correlation energy density. This approximation is called the local density approximation (LDA). The usual approach to construct E^{XC} in the LDA is to take the density to be locally uniform, so that one can locally use the exchange-correlation energy of a uniform electron gas with the density of the real system at a given point [4, 5, 45]. Taking into account that correlation effects are non-local and that the density in a real crystal is far from being uniform, the thus-constructed LDA seems to be too drastic an approximation. However, extensive applications have proven such an approach to be very successful, the reasons for this success being not completely clear.

For the uniform electron gas the exchange energy density is known exactly [4, 46, 47],

$$\epsilon^{X}(n) = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{\frac{1}{3}} n^{\frac{1}{3}} = -\frac{3}{4}\left(\frac{9}{4\pi^2}\right)^{\frac{1}{3}} \frac{1}{r_{\mathrm{s}}}, \qquad (2.47)$$

where $r_{\mathrm{s}} = (3/4\pi n)^{1/3}$ is the radius of the sphere that an electron would on average occupy in the uniform electron gas of density n. For the correlation energy density, analytic expressions exist only in the high [48] and low density limits. Accurate Monte Carlo results [49] are used for the intermediate densities. Usually differently parametrized analytic forms of ϵ^{XC} are used to interpolate between these known results. In the calculations that we present in this thesis, we used the parametrization suggested by Goedecker, Teter, and Hutter [50], namely

$$\epsilon^{\mathrm{XC}}(r_{\mathrm{s}}) = -\frac{A_0 + A_1 r_{\mathrm{s}} + A_2 r_{\mathrm{s}}^2 + A_3 r_{\mathrm{s}}^3}{B_1 r_{\mathrm{s}} + B_2 r_{\mathrm{s}}^2 + B_3 r_{\mathrm{s}}^3 + B_4 r_{\mathrm{s}}^4}. \qquad (2.48)$$

The parameters A_i and B_i that we used are those listed in Ref. [50].

2.2.4 Pseudopotentials: simplifying atomic potentials

As mentioned above, it is mainly the valence electrons of atoms that are involved in the formation of a solid, while electrons in the core shells almost do not feel the crystalline environment. For this reason, it is desirable to take the core electrons out of consideration. This can be done by means of constructing a pseudopotential [51] – a potential that is much smoother compared to the real ionic potential but has (ideally) the same effect on the valence electrons.

Consider an isolated atom, described by a single-particle Hamiltonian $\tilde{H}_{\mathrm{sp}}$, where tilde distinguishes the atomic Hamiltonian from the crystalline one. The construction of such a Hamiltonian (and the corresponding effective potential $\tilde{V}_{\mathrm{eff}}$) for an isolated atom is completely analogous to the case of a crystal, discussed in the preceding sections. Assume the single-particle energies and wavefunctions of the core $|\phi_n^{(c)}\rangle$ and valence $|\phi_\alpha^{(v)}\rangle$ electrons to be known [2]:

$$\begin{aligned}\tilde{H}_{\mathrm{sp}}|\phi_n^{(\mathrm{c})}\rangle &= \epsilon_n^{(\mathrm{c})}|\phi_n^{(\mathrm{c})}\rangle,\\ \tilde{H}_{\mathrm{sp}}|\phi_\alpha^{(\mathrm{v})}\rangle &= \epsilon_\alpha^{(\mathrm{v})}|\phi_\alpha^{(\mathrm{v})}\rangle.\end{aligned} \tag{2.49}$$

The trickery of the pseudopotential method consists in the introduction of a new set of states $|\bar{\phi}_\alpha^{(\mathrm{v})}\rangle$ that are the eigenstates of some different single-particle Hamiltonian, but with the same eigenvalues as the original valence states, and that are also much smoother in the core region.

These states can be defined in many ways. For example, following Phillips and Kleinman [51], one can define them as a special superposition of the core and valence states of the original problem

$$|\phi_\alpha^{(\mathrm{v})}\rangle = |\bar{\phi}_\alpha^{(\mathrm{v})}\rangle - \sum_n \langle\phi_n^{(\mathrm{c})}|\bar{\phi}_\alpha^{(\mathrm{v})}\rangle|\phi_n^{(\mathrm{c})}\rangle, \tag{2.50}$$

where the summation is over all the core states, and one can recognize the projector onto the core states $P_{\rm c} = \sum_n |\phi_n^{\rm (c)}\rangle\langle\phi_n^{\rm (c)}|$. Plugging this expression for $|\phi_\alpha^{\rm (v)}\rangle$ into the first equation of (2.49), one arrives to the eigenvalue equation for $\bar{\psi}$

$$\left[\tilde{H}_{\rm sp} + \sum_n (\epsilon_\alpha^{\rm (v)} - \epsilon_n^{\rm (c)})|\phi_n^{\rm (c)}\rangle\langle\phi_n^{\rm (c)}|\right]|\bar{\psi}_\alpha^{\rm (v)}\rangle = \epsilon_\alpha^{\rm (v)}|\bar{\psi}_\alpha^{\rm (v)}\rangle. \tag{2.51}$$

We see that we ended up with a single-particle equation with a modified effective potential

$$\tilde{V}^{\rm PS} = \tilde{V}_{\rm eff} + \sum_n (\epsilon_\alpha^{\rm (v)} - \epsilon_n^{\rm (c)})|\phi_n^{\rm (c)}\rangle\langle\phi_n^{\rm (c)}|, \tag{2.52}$$

which results in the same spectrum as that of the valence electrons in the original problem. The core electrons are now completely eliminated.

It can immediately be seen from the above expression that $\tilde{V}^{\rm PS}$ acts differently on the states of different angular momentum. Although the above described scheme of constructing a pseudopotential is far from being unique, the angular momentum dependence is a general feature – in order to give the correct scattering properties of the original atomic potential, the pseudopotential should be angular-momentum dependent. Thus, a more general form of a pseudopotential is

$$\tilde{V}^{\rm PS}(\mathbf{r}) = \sum_{\ell=0}^{\infty}\sum_{m=-\ell}^{\ell} v_\ell^{\rm PS}(r)|\ell m\rangle\langle\ell m|, \tag{2.53}$$

where $|\ell m\rangle$ are spherical harmonics. If $v_\ell^{\rm PS}$ are the same for every ℓ, the pseudopotential is said to be local. It is in principle possible to reproduce the correct phase shifts for each angular momentum component of the valence wavefunction with a local potential [4]. However, most of the existing pseudopotentials have a non-local[1] part, which generally makes them much smoother than their local counterparts.

[1] Since in the radial coordinate $v_\ell^{\rm PS}(\mathbf{r})$ are local, the term "semi-local" is sometimes used instead.

The components v_ℓ^{PS} are chosen such that for each ℓ the ground state of the pseudopotential is the valence ℓ-state. For example, in the case of silicon, the $3s$ and $3p$ valence state of the original atomic potential should become the $1s$ and $2p$ ground states of $\ell = 0$ and $\ell = 1$ pseudopotentials correspondingly. Another desirable property of a pseudopotential is its transferability, i.e., it should not depend on the particular atomic environment.

A powerful scheme for constructing transferable pseudopotentials was proposed by Hamann, Schlüter and Chiang [52]. They introduced the following four constraints on the pseudo-wavefunction $\bar{\psi}^{(\mathrm{v})}$:

(1) the pseudo-wavefunction and the original all-electron wavefunction should correspond to the same eigenvalue;

(2) the radial part of a pseudo-wavefunction, R_{PS}, has to be nodeless and coincide with that of the all-electron wavefunction, R_{AE}, outside a sphere of some *a priori* chosen radius r_{c};

(3) the norm-conservation of the pseudo-wavefunction should be obeyed for $r < r_{\mathrm{c}}$:

$$\int_0^{r_c} r^2 |R_{\mathrm{PS}}(r)|^2 dr = \int_0^{r_c} r^2 |R_{\mathrm{AE}}(r)|^2 dr; \qquad (2.54)$$

(4) the logarithmic derivatives of R_{PS} and R_{AE} should agree for $r \geq r_{\mathrm{c}}$.

Given a wavefunction that satisfies the above conditions, the Schrödinger equation for its radial part is inverted to give the corresponding *norm-conserving* pseudopotential. There are several very reliable sets of norm-conserving pseudopotentials available [53–55], including the ones of Goedecker, Teter and Hutter (GTH) that have a simple analytic form [50].

The use of pseudopotentials in calculations on solids allows one to replace a steep ionic potential with a smooth potential that accounts for screening of the nucleus by core electrons. Thus, if the pseudopotential approach is used, the resultant V_{eff} of Eq. (2.43) is also smooth. Smoothness allows for the use of a plane wave basis set in the calculations, which is very convenient. The smoother is the pseudopotential, the smoother are the resultant single-particle wavefunctions, and the less plane waves are needed to fit them. The use of plane waves with pseudopotentials is thus the basis of a powerful numerical technique.

There are many excellent numerical packages based on this technique, including some that are freely available. We used one such code – Abinit [56, 57] – for the purposes of the present work. The wavefunction in the plane wave basis has the form given by the decomposition of Eq. (2.25), and the code stores **G** and $C_n(\mathbf{G})$ for each state (of the afore-specified set) at each **k**-point of the mesh.

Alas, the number of plane waves used in the decomposition at different **k**-points is different, and, in general, involves different sets of **G**-vectors. In this respect inner products, like the overlap of the form $\langle u_{n\mathbf{k}}|u_{m\mathbf{k}+\Delta\mathbf{k}}\rangle$, might seem to be ill-defined. However, when $\Delta\mathbf{k}$ is small (that is, when the **k**-mesh is dense), then for those vectors in the set $\{\mathbf{G}\}$ at $\mathbf{k}$ that are not present in the set at $\mathbf{k}+\Delta\mathbf{k}$ (and vice versa) the corresponding coefficients $C(\mathbf{G})$ are small, so that one can safely approximate the above overlap as

$$\langle u_{n\mathbf{k}}|u_{m\mathbf{k}+\Delta\mathbf{k}}\rangle = {\sum_{\mathbf{G}}}' C^*_{n\mathbf{k}}(\mathbf{G}) C_{m\mathbf{k}+\Delta\mathbf{k}}(\mathbf{G}), \tag{2.55}$$

where prime means that the summation goes over those **G**-vectors that are present in the decomposition of both $u_{n\mathbf{k}}$ and $u_{m\mathbf{k}+\Delta\mathbf{k}}$.

2.2.5 Inclusion of spin-orbit interaction

In the following we are interested in applications of DFT with spin-orbit interaction taken into account. We now briefly discuss how this is implemented.

Spin-orbit coupling arises from relativistic effects deep in the core. Thus, it can be included into the pseudopotential. As mentioned above, in the non-relativistic case the most general form of a pseudopotential is that of Eq. (2.53).

In the relativistic case ℓ is not a good quantum number anymore, and the total angular momentum $\mathbf{J} = \mathbf{L} + \mathbf{S}$ is conserved instead. For electrons with spin 1/2 this results in two values $j = \ell \pm 1/2$ of the quantum number j for each orbital angular momentum ℓ. As suggested in Ref. [55], it is sufficient to construct a pseudopotential for both values of j using the relativistic all-electron calculations. Then the scalar-relativistic part of the pseudopotential is given by the weighted average of the two terms

$$v_\ell^{\mathrm{PS}} = \frac{1}{2\ell+1}\left[(\ell+1)v_{\ell+1/2}^{\mathrm{PS}} + \ell v_{\ell-1/2}^{\mathrm{PS}}\right], \tag{2.56}$$

while spin-orbit part is captured by

$$\delta v_\ell^{\mathrm{SO}} = \frac{2}{2\ell+1}\left[v_{\ell+1/2}^{\mathrm{PS}} - v_{\ell-1/2}^{\mathrm{PS}}\right], \tag{2.57}$$

and gives a contribution to the pseudopotential of the form

$$V_{\mathrm{SO}}^{\mathrm{PS}} = \sum_\ell \sum_m |\ell m\rangle \delta v_\ell^{\mathrm{SO}} \mathbf{L}\cdot\mathbf{S}\langle \ell m|. \tag{2.58}$$

Then the total pseudopotential

$$V_{\mathrm{total}}^{\mathrm{PS}} = V^{\mathrm{PS}} + V_{\mathrm{SO}}^{\mathrm{PS}} \tag{2.59}$$

replaces the electron-ion term in the corresponding single-particle equation. As discussed above, this potential is smooth compared to the original ionic potential and allows one to use plane wave decompositions. The spinor wavefunction has the form

$$\psi_{n\mathbf{k}}(\mathbf{r}) = \begin{pmatrix} \varphi_{n\uparrow\mathbf{k}}(\mathbf{r}) \\ \varphi_{n\downarrow\mathbf{k}}(\mathbf{r}) \end{pmatrix} \tag{2.60}$$

with both components φ written in the form of Eq. (2.25), and with the normalization condition

$$\int d\mathbf{r} \left(|\varphi_{n\uparrow\mathbf{k}}(\mathbf{r})|^2 + |\varphi_{n\downarrow\mathbf{k}}(\mathbf{r})|^2 \right) = 1. \tag{2.61}$$

For spinor wavefunctions, overlaps like that in Eq. (2.55) are computed for each of the spinor components separately.

One of the possibilities to include spin-orbit interaction in Abinit calculation is to use the fully relativistic pseudopotentials of Harwigsen, Goedecker and Hutter [58] (HGH). These pseudopotentials start with an analytic form of the above mentioned GTH pseudopotentials and then spin-orbit coupling is included according the recipe of this subsection.

2.3 Tight-binding approximation

We now discuss a more qualitative method widely used in band-structure calculations, namely the tight-binding approximation (TBA). The main idea of the method is to take into account only those atomic orbitals that are responsible for the physical effect in question. These orbitals are then used to construct Bloch-like functions that serve as a basis for a tight-binding Hamiltonian.

Let us consider a unit cell with ℓ atoms in it. We use $\mathbf{R}$ to label the lattice vector, $\mathbf{t}_j$ to label the position of the j-th atom in the home unit cell ($\mathbf{R} = \mathbf{0}$), and $\bar{s}$ to label basis orbitals (not necessarily atomic). Let us also introduce a label $\tau = \{\bar{s}, j\}$ that describes a given orbital $\bar{s}$ on a given atom j. In the case of a single basis orbital, the label $\bar{s}$ will be omitted. In these notations the orbital wavefunction is $\phi_\tau(\mathbf{r}) = \phi_{\bar{s}}(\mathbf{r} - \mathbf{t}_j) = \langle \mathbf{r} | \mathbf{0} \bar{s} j \rangle = \langle \mathbf{r} | \mathbf{0} \tau \rangle$, where the label $\mathbf{0}$ refers to the lattice vector, i.e., this orbital belongs to the home unit cell. To obtain the orbitals in the other cells, the translation operator is used:

$$\hat{T}_{\mathbf{R}} \phi_\tau(\mathbf{r}) = \phi_\tau(\mathbf{r} - \mathbf{R}) = \langle \mathbf{r} | \mathbf{R} \tau \rangle. \tag{2.62}$$

We now use these basis orbitals to construct Bloch-like functions

$$\chi_{\mathbf{k}\tau}(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \phi_\tau(\mathbf{r} - \mathbf{R}) \tag{2.63}$$

or in the Dirac notations

$$|\chi_{\mathbf{k}\tau}\rangle = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} |\mathbf{R}\tau\rangle. \tag{2.64}$$

The functions are Bloch-like in the sense that

$$\chi_{\mathbf{k}\tau}(\mathbf{r} + \mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}} \chi_{\mathbf{k}\tau}(\mathbf{r}), \tag{2.65}$$

which is easy to check. We look for the eigenstates of the single-particle Hamiltonian in the form

$$|\psi_{n\mathbf{k}}\rangle = \sum_{\tau} C_{\tau n\mathbf{k}} |\chi_{\mathbf{k}\tau}\rangle. \tag{2.66}$$

Thus, the eigenstate of $H_{\mathbf{k}}$ is represented by a column vector of coefficients $C_{\tau n\mathbf{k}}$.

The point of the usual application of the TBA is to get a qualitative understanding of some effect. The method can be made quantitative (to some extent) by, for example, matching the Hamiltonian parameters to those obtained from DFT. This is possible, since the concept of the TBA is to construct $H_{\mathbf{k}}$ by explicitly specifying the effect of the single-particle Hamiltonian on the basis states. For this purpose the Hamiltonian is written in a tight-binding basis

$$H_{\rho\tau}(\mathbf{k}) = \langle\chi_{\mathbf{k}\rho}|H_{\text{sp}}|\chi_{\mathbf{k}\tau}\rangle = \frac{1}{N} \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \langle\mathbf{0}\rho|H_{\text{sp}}|\mathbf{R}\tau\rangle. \tag{2.67}$$

If L is the number of orbitals per atom times the number of atoms in the unit cell, then $H(\mathbf{k})$ is an $L \times L$ matrix. The other matrix that plays a crucial role is the $L \times L$ overlap matrix

$$S_{\rho\tau}(\mathbf{k}) = \langle\chi_{\mathbf{k}\rho}|\chi_{\mathbf{k}\tau}\rangle \tag{2.68}$$

Now we are in a position to write down the Schrödinger equation

$$H_{\mathrm{sp}}|\psi_{n\mathbf{k}}\rangle = \epsilon_{n\mathbf{k}}|\psi_{n\mathbf{k}}\rangle$$

in the TBA. Using the ansatz of Eq. (2.66) and multiplying by a bra vector $\langle\chi_{\mathbf{k}\rho}|$ on the left we arrive at

$$\sum_{\tau} C_{\tau n\mathbf{k}}\langle\chi_{\mathbf{k}\rho}|H_{\mathrm{sp}}|\chi_{\mathbf{k}\tau}\rangle = \sum_{\tau} C_{\tau n\mathbf{k}}\langle\chi_{\mathbf{k}\rho}|\chi_{\mathbf{k}\tau}\rangle\epsilon_{n\mathbf{k}} \tag{2.69}$$

or, in the above notations

$$\sum_{\tau} H_{\rho\tau}(\mathbf{k})C_{\tau n\mathbf{k}} = \epsilon_{n\mathbf{k}}\sum_{\tau} S_{\rho\tau}(\mathbf{k})C_{\tau n\mathbf{k}}. \tag{2.70}$$

Finally, this result can be written in the matrix form

$$H(\mathbf{k})C_{n\mathbf{k}} = \epsilon_{n\mathbf{k}}S(\mathbf{k})C_{n\mathbf{k}}, \tag{2.71}$$

where $H(\mathbf{k})$ and $S(\mathbf{k})$ are $L \times L$ matrices and $C_{n\mathbf{k}}$ are L-component column vectors. This is a generalized eigenvalue problem, and the energy eigenvalues ϵ_{nk} are found from the equation

$$\det\left[H(\mathbf{k}) - \epsilon_{n\mathbf{k}}S(\mathbf{k})\right] = 0. \tag{2.72}$$

For the qualitative calculations in the TBA that we are going to use, it is sufficient to consider an orthogonal basis

$$\langle\mathbf{0}i|\mathbf{R}j\rangle = \int \phi_i^*(\mathbf{r})\phi_j(\mathbf{r})d\mathbf{r} = \delta_{ij}, \tag{2.73}$$

which leads to $S = I$, where I is the unit matrix. In general, in this approximation $\langle\mathbf{0}\rho|H_{\mathrm{sp}}|\mathbf{R}\tau\rangle \neq 0$, while $\langle\mathbf{0}\rho|\mathbf{R}\tau\rangle = \delta(\mathbf{R})\delta_{\rho\tau}$, which is not very realistic, but it turns out to be very useful as a first iteration to the problem. In Chapter 4 we will see that under these conditions the states $|\mathbf{R}\tau\rangle$ are closely related to Wannier functions.

In the following chapters we will use this scheme to construct simple models of topological insulators, that will have all the features of the non-trivial topological band structure captured within the minimal possible occupied space.

Overlaps between the cell-periodic Bloch states

Before we proceed, we briefly comment on the calculation of overlaps of the form $M_{nm}^{(\mathbf{k},\mathbf{k}+\Delta\mathbf{k})} = \langle u_{n\mathbf{k}}|u_{m\mathbf{k}+\Delta\mathbf{k}}\rangle$ in the TBA, since it is one of the main ingredients of many calculations in the following chapters.

In our notations $|\chi_{\mathbf{k}+\mathbf{G},\tau}\rangle = |\chi_{\mathbf{k}\tau}\rangle$, and in order to have $\psi_{n,\mathbf{k}+\mathbf{G}} = \psi_{n\mathbf{k}}$ the coefficients must obey

$$C_{\tau n,\mathbf{k}+\mathbf{G}} = C_{\tau n\mathbf{k}}. \tag{2.74}$$

For the periodic parts of the Bloch states $u_{n\mathbf{k}}$ we have the usual relation

$$u_{n\mathbf{k}}(\mathbf{r}) = \frac{e^{-i\mathbf{k}\cdot\mathbf{r}}}{\sqrt{N}} \sum_{\tau} C_{\tau n\mathbf{k}} \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot(\mathbf{R}+\mathbf{r})}\phi_{\tau}(\mathbf{r}-\mathbf{R}) = e^{-i\mathbf{k}\cdot\mathbf{r}}\psi_{n\mathbf{k}}(\mathbf{r}). \tag{2.75}$$

The explicit calculation of overlaps gives

$$M_{mn}^{(\mathbf{k},\mathbf{k}+\Delta\mathbf{k})} = \frac{1}{N}\int d\mathbf{r} \left[\sum_{\tau\rho} C^{*}_{\tau m\mathbf{k}} C_{\rho n,\mathbf{k}+\Delta\mathbf{k}} \sum_{\mathbf{R},\mathbf{R}'} e^{i\mathbf{k}\cdot(\mathbf{R}'-\mathbf{R})}\phi^{*}_{\tau}(\mathbf{r}-\mathbf{R})\phi_{\rho}(\mathbf{r}-\mathbf{R}')e^{i\Delta\mathbf{k}\cdot(\mathbf{R}'-\mathbf{r})}\right] \tag{2.76}$$

As discussed above, we assume the atomic orbitals to be strongly localized on the atomic sites and mutually orthogonal, $\langle\mathbf{R}i|\mathbf{R}'j\rangle = \delta_{ij}\delta(\mathbf{R}-\mathbf{R}')$, and the overlap integral is non-zero only if the two orbitals are actually the same.

In the applications of the TBA in the following chapters, we consider an extreme limit of δ-like orbitals, peaked at the given lattice cite. In that case

$\phi_\tau(\mathbf{r}-\mathbf{R}) \sim \delta(\mathbf{r}-\mathbf{R}-\mathbf{t}_j)$, and we get

$$M_{mn}^{(\mathbf{k},\mathbf{k}+\Delta\mathbf{k})} = \sum_{\tau=\{\bar{s},j\}} C^*_{\tau m\mathbf{k}} C_{\tau n\mathbf{k}+\Delta\mathbf{k}} e^{-i\Delta\mathbf{k}\cdot\mathbf{t}_j}. \tag{2.77}$$

Chapter 3
Topological insulators

In recent years the band theory of solids has been augmented by new chapters to account for geometric and topological effects that had not been considered previously. The introduction of the Berry phase [6] allowed the systematic description of many observable effects of purely geometric origin, such as the Aharonov-Bohm effect [59], and its applications in the band-theory context have included the theory of electric polarization [10, 11] and the anomalous Hall conductance [12, 13].

More recently it was realized that insulating band structures can be topologically different [1, 19, 24]. On an intuitive level a topological distinction is understood as the impossibility of a smooth transformation of one object into another. For example, it is impossible to smoothly transform a sphere into a torus without pinching a hole in it. This means that a sphere and a torus belong to different topological classes. On the other hand, a smooth deformation of a sphere into a cube is obviously possible, so that a cube and a sphere are in the same topological class.

However, it is not always easy to say whether a smooth deformation of one geometric shape into another exists, especially in higher dimensions, and some rigorous existence criterion should be developed. For the case of two-dimensional surfaces such a criterion is given by the Gauss-Bonnet theorem that relates the surface integral of the Gaussian curvature K taken over the surface[1] of the geometric object M to an integer g called the genus – the number of holes in the

[1] Here for simplicity we assume that the surface has no boundary.

object [60]

$$\int_M K dA = 2 - 2g, \tag{3.1}$$

thus classifying surfaces according to the value of the topological invariant g.

We will see that insulating band structures are classified in an analogous way, with the Gaussian curvature replaced by other geometric quantities. In the spirit of the above example of geometric shapes, gapped band structures that belong to different topological classes cannot be continuously deformed into one another. A continuous deformation in this case is done by means of adiabatic changes in the Hamiltonian, and the impossibility of going from one class to another is reflected in the fact that different topological insulating phases are separated by a metallic phase, where the band gap closes.

The reason for a topological distinction between insulators is however somewhat different than in the case of the geometric example considered above. As mentioned in section 2.1.2 , the BZ can be considered as a closed manifold. For an insulator there is a vector space of occupied states with dimensionality $\mathcal{N}$ attached to every point of this manifold. That is, at each $\mathbf{k}$-point in the BZ the Hamiltonian has $\mathcal{N}$ occupied valence states, separated by an energy gap from the conduction states. The topological classification arises from the topologically different ways in which these vector spaces can be glued together on the whole BZ manifold.

To understand this point, consider a simple example of a circle with a segment $[-1, 1]$ attached to each of its points. One can imagine different scenarios here. First, imagine that the segments are oriented in the same way at each point of the circle; in this case we end up with a circular ribbon. Now, imagine that the segments rotate around the circular base in a continuous manner, so that the total rotation angle is π. In this case we end up with a Möbius strip, which is obviously topologically distinct from the ribbon. By analogy, different "gluing" of vector spaces on the BZ manifold results in different topologies of insulating band structures.

With this intuitive understanding of the nature of topological phases in insulators, let us now put the problem on a more rigorous footing. We will first discuss certain geometric quantities and their appearance in band theory, and then discuss two topological insulating phases known so far: $\mathcal{T}$-breaking Chern insulators and $\mathcal{T}$-symmetric $\mathbb{Z}_2$ insulators.

3.1 Gauge transformations

In what follows we consider a single-particle Hamiltonian with V_{eff} replacing the many-body terms, and consider the manifold of occupied single-particle states of this Hamiltonian.

It is known from a basic course in quantum mechanics that if a state vector is multiplied by a phase factor,

$$|\tilde{\psi}\rangle = e^{i\varphi}|\psi\rangle, \tag{3.2}$$

no observables are changed and this new state describes the same physics as the old one. In band structure, the above relation obviously holds for single-particle states $|\psi\rangle \equiv |\psi_{n\mathbf{k}}\rangle$, but actually the freedom in choosing particular single-particle states to describe a solid becomes much broader. Due to the fact that the crystal ground state is a many-particle state, hidden behind the single-particle description (i.e., the ground state is given by a collection of occupied single-particle states), the whole set of occupied states contributes to observables. Thus, instead of changing a U(1) phase of individual single-particle states, it makes sense to consider more general transformations.

For an insulator in its ground state, all the observables are uniquely defined by the projector onto the occupied space

$$\hat{P}_{\mathrm{o}}(\mathbf{k}) = \sum_{n=1}^{\mathcal{N}} |\psi_{n\mathbf{k}}\rangle\langle\psi_{n\mathbf{k}}|, \tag{3.3}$$

where $\mathcal{N}$ is the number of occupied bands. The average value of any observable $\hat{\mathcal{O}}$ is given by the trace, taken in the occupied Hilbert space:

$$< \hat{\mathcal{O}} >_{\mathbf{k}} = \mathrm{Tr}\left[\hat{P}_{\mathrm{o}}\hat{\mathcal{O}}\right] = \sum_{n=1}^{\mathcal{N}} \langle\psi_{n\mathbf{k}}|\hat{\mathcal{O}}|\psi_{n\mathbf{k}}\rangle. \tag{3.4}$$

Since the trace of a matrix is invariant with respect to unitary transformations, it follows from the above expression that any $\mathcal{N}$ orthonormal vectors that span the occupied Hilbert space can be taken to describe the insulating ground state. In particular, starting with the Hamiltonian eigenstates, a **k**-dependent unitary transformation $\mathcal{U}(\mathbf{k}) \in \mathrm{U}(\mathcal{N})$ of the states can be carried out to get a new set of states

$$|\tilde{\psi}_{n\mathbf{k}}\rangle = \sum_{m=1}^{\mathcal{N}} \mathcal{U}_{mn}(\mathbf{k})|\psi_{m\mathbf{k}}\rangle \tag{3.5}$$

that gives an equally good description of the insulator. It is important to note that the states $\tilde{\psi}_{n\mathbf{k}}$ are not necessarily Hamiltonian eigenstates anymore. The choice of particular representatives for the occupied Hilbert space is referred to as a gauge choice, in analogy with electromagnetism, where different gauge choices for the electromagnetic potential can be used to describe the same physical phenomena. A particular gauge choice is a matter of convenience.

As is the case for normal observables, the topological class of an insulator is also determined by the whole occupied space. Thus, a gauge choice does not have any effect on the topology of a gapped band structure, while individual single-particle states of the occupied space might acquire some non-trivial phases. This issue is discussed in more detail in Chapter 7. For now, let us just mention that despite the topological index of an insulating phase being a gauge-independent quantity, particular expressions used to calculate it might contain gauge-dependent terms, as shown below. In principle, for any gauge-independent quantity, a gauge-independent expression should exist, but it is not always easy to find it.

3.2 Geometric phases in band theory

Until the middle of the 1980's the phase of a quantum state accumulated in the process of adiabatic evolution was typically disregarded. Fock presented an argument [61] that this phase can always be taken to be unity. However, this result was derived with the assumption of non-cyclic evolution. Surprisingly, a general theory of cyclic evolutions was not considered until 1984, when Berry, in his seminal paper of Ref. [6], proved that a cyclic evolution of a quantum state results in a phase factor of a purely geometric origin and, in principle, is observable. Nowadays the Berry phase is ubiquitous in physics and its discussion is included in most of the contemporary textbooks on quantum mechanics. For completeness, we provide a quick review here.

3.2.1 Berry phase

Consider a system that depends on some external parameter $\boldsymbol{\xi}$. The parameter may vary with time, so in the most general case we work with a Hamiltonian $H(\boldsymbol{\xi}(t))$. The quantum adiabatic theorem [62] states that a system initially in a Hamiltonian eigenstate $|\psi_n(0)\rangle = |n(\boldsymbol{\xi}(0))\rangle$ will remain in the instantaneous eigenstate of the Hamiltonian in the process of adiabatic evolution:

$$H(\boldsymbol{\xi}(t))|n(\boldsymbol{\xi}(t))\rangle = E_n(\boldsymbol{\xi}(t))|n(\boldsymbol{\xi}(t))\rangle \tag{3.6}$$

However, apart from the usual dynamical phase, the state can acquire an extra phase factor

$$|\psi_n(t)\rangle = e^{i\gamma_n(t)} e^{-i/\hbar \int_0^t \epsilon_n(\boldsymbol{\xi}(t'))dt'} |n(\boldsymbol{\xi}(t))\rangle. \tag{3.7}$$

Plugging this state into the time-dependent Schrödinger equation with the Hamiltonian $H(\boldsymbol{\xi}(t))$, and multiplying the resultant equation on the left with $\langle n(\boldsymbol{\xi}(t))|$, the extra phase is expressed as

$$\gamma_n = \int_{\mathcal{C}} \boldsymbol{\mathcal{A}}_n(\boldsymbol{\xi}) \cdot d\boldsymbol{\xi}, \tag{3.8}$$

where

$$\boldsymbol{\mathcal{A}}_n(\boldsymbol{\xi}) = i\langle n(\boldsymbol{\xi})| \frac{\partial}{\partial \boldsymbol{\xi}} |n(\boldsymbol{\xi})\rangle \tag{3.9}$$

and $\mathcal{C}$ is the contour traversed by the adiabatic parameter $\boldsymbol{\xi}$ during the evolution. The connection $\boldsymbol{\mathcal{A}}$ is obviously gauge-dependent. A gauge transformation

$$|n'(\boldsymbol{\xi})\rangle = e^{i\varphi(\boldsymbol{\xi})} |n(\boldsymbol{\xi})\rangle \tag{3.10}$$

changes it to

$$\boldsymbol{\mathcal{A}}_n' = \boldsymbol{\mathcal{A}}_n - \frac{\partial}{\partial \boldsymbol{\xi}} \varphi(\boldsymbol{\xi}). \tag{3.11}$$

Hence, the phase is changed by $\varphi(\boldsymbol{\xi}(0)) - \varphi(\boldsymbol{\xi}(t_\mathrm{f}))$, where $\boldsymbol{\xi}(0)$ and $\boldsymbol{\xi}(t_\mathrm{f})$ are the endpoints of the contour $\mathcal{C}$. A suitable choice of φ thus makes γ_n vanish [61] when $\boldsymbol{\xi}(0) \neq \boldsymbol{\xi}(t_\mathrm{f})$.

However, things change once one considers cyclic evolution with $\boldsymbol{\xi}(0) = \boldsymbol{\xi}(t_\mathrm{f})$. Since the wavefunction in Eq. (3.10) has to be singlevalued, we have

$$\varphi(\boldsymbol{\xi}(0)) - \varphi(\boldsymbol{\xi}(t_\mathrm{f})) = 2\pi m, \tag{3.12}$$

where $m \in \mathbb{Z}$. Thus, γ_n can not be removed anymore, and the Berry phase

$$\gamma_n = \oint_C \boldsymbol{\mathcal{A}}_n(\boldsymbol{\xi}) \cdot d\boldsymbol{\xi}, \tag{3.13}$$

becomes well-defined. As soon as the adiabaticity requirement is satisfied, γ_n is insensitive to a particular form of the time dependence of $\boldsymbol{\xi}$, and for this reason we left time out of the formulas.

3.2.2 Applications in band theory

In band theory the Bloch states have an explicit **k**-dependence, which gives a natural parameter to study Berry phase effects [7]. We first consider an isolated band – a band that is separated by energy gaps from the rest of the spectrum. Later, we generalize our consideration to a many-band case.

For a given isolated band n one can define an (Abelian) Berry connection in accord with Eq. (3.9). Since we are interested in a three-dimensional Euclidean space, $\frac{\partial}{\partial \boldsymbol{\xi}}$ is replaced with $\nabla_{\mathbf{k}}$ to give

$$\boldsymbol{\mathcal{A}}_n(\mathbf{k}) = i\langle u_{n\mathbf{k}}|\nabla_{\mathbf{k}}|u_{n\mathbf{k}}\rangle. \tag{3.14}$$

Here we use $u_{n\mathbf{k}}$ and not $\psi_{n\mathbf{k}}$ in order for the **k**-derivative to be well-defined, as discussed in Chapter 2. We have seen above that the Berry connection is gauge-dependent. In this respect, $\boldsymbol{\mathcal{A}}$ is analogous to the electromagnetic potential, and often the term “gauge potential” is used in the literature. Continuing along this line, a “gauge field” or Berry curvature is defined in analogy with the magnetic field:

$$\boldsymbol{\mathcal{F}}(\mathbf{k}) = \nabla_{\mathbf{k}} \times \boldsymbol{\mathcal{A}}(\mathbf{k}). \tag{3.15}$$

Later we will see that it indeed acts like a magnetic field in **k**-space. It is easy to check that the Berry curvature does not change under a U(1) gauge transformation.

Let us now consider $\mathcal{N} > 1$ bands separated by energy gaps from the rest of the spectrum. The Abelian connection of Eq. (3.14) is now replaced by its non-Abelian multiband generalization [63, 64]

$$\mathcal{A}_{mn,\alpha} = i\langle u_{m\mathbf{k}}|\partial_{k_\alpha}|u_{n\mathbf{k}}\rangle \tag{3.16}$$

and the non-Abelian curvature is defined as

$$F_{mn,\alpha\beta} = \mathcal{F}_{mn,\alpha\beta} - i[\mathcal{A}_\alpha, \mathcal{A}_\beta]_{mn}, \tag{3.17}$$

where the $\mathbf{k}$-dependence is implicit, and

$$\mathcal{F}_{mn,\alpha\beta} = i\left[\langle\partial_{k_\alpha}u_{n\mathbf{k}}|\partial_{k_\beta}u_{m\mathbf{k}}\rangle - \langle\partial_{k_\beta}u_{n\mathbf{k}}|\partial_{k_\alpha}u_{m\mathbf{k}}\rangle\right].$$

The curvature F is gauge-covariant and $\mathrm{Tr}[\mathbf{F}_{\alpha\beta}] = \mathrm{Tr}[\boldsymbol{\mathcal{F}}_{\alpha\beta}]$ is gauge-invariant [18] under a general unitary transformation $\mathcal{U} \in \mathrm{U}(\mathcal{N})$ of Eq. (3.5).

Analogous to the case of Gaussian curvature, an integral of the Berry curvature over a closed manifold is an integer-valued topological invariant [18] called a Chern number or, more specifically, a first Chern number. Since a BZ is a closed compact manifold with no boundary, integration of $\mathcal{F}(\mathbf{k})$ over the BZ should give an integer. According to this, in two dimensions for the Chern number we have

$$C = \frac{1}{2\pi}\iint_{BZ} \mathrm{Tr}\left[F_{\alpha\beta}\right] dk_\alpha dk_\beta. \tag{3.18}$$

In three dimensions, one can define three Chern numbers $\{C_\alpha, C_\beta, C_\gamma\}$. Some explanation is in order here. Since for a given value of k_α a cross section of the BZ that is orthogonal to the direction $\hat{k}_\alpha$ forms an effectively two-dimensional

BZ, the above equation can be used to define $C_\alpha(k_\alpha)$ for this given k_α. However, since $C_\alpha(k_\alpha)$ has to be an integer, it cannot change smoothly, and for this reason, a smooth change in k_α should result in $C_\alpha(k_\alpha + \delta k_\alpha) = C_\alpha(k_\alpha)$ as long as the set of bands for which C_α is defined remains isolated when going from k_α to $k_\alpha + \delta k_\alpha$. Thus, in an insulator, C_α of a whole occupied space is a well-defined topological invariant.

3.2.3 Numerical computation of Berry curvature

There are several ways to compute the above-defined quantities numerically. When possible, it is preferable to work with the gauge-invariant curvature, and not the connection, which is gauge-dependent. Numerical diagonalization brings random phases to the states at different **k**-points on the mesh (random gauge), and computing the connection requires additional gauge fixing.

When dealing with a single isolated band, one can express the curvature as

$$\mathcal{F}_{n,\alpha\beta} = -2\mathrm{Im}\sum_{m=1}^{\mathcal{N}}{}' \frac{\langle u_{n\mathbf{k}}|\partial_{k_\alpha} H(\mathbf{k})|u_{m\mathbf{k}}\rangle\langle u_{m\mathbf{k}}|\partial_{k_\beta} H(\mathbf{k})|u_{n\mathbf{k}}\rangle}{(\epsilon_{n\mathbf{k}} - \epsilon_{m\mathbf{k}})^2} \tag{3.19}$$

where the prime means that the term $m = n$ is omitted in the summation. This expression can be easily derived [9] using perturbation a expansion for the wavefunction $u_{n,\mathbf{k}+\Delta\mathbf{k}}$. It leads to the expression for the gradient

$$\nabla_{\mathbf{k}}|u_{n\mathbf{k}}\rangle = \sum_{m\neq n} \frac{\langle u_{n\mathbf{k}}|\nabla_{\mathbf{k}} H(\mathbf{k})|u_{n\mathbf{k}}\rangle}{\epsilon_{n\mathbf{k}} - \epsilon_{m\mathbf{k}}}|u_{m\mathbf{k}}\rangle, \tag{3.20}$$

from which the expression (3.19) follows.

We now consider an alternative approach, which also applies to the multiband case. Here one computes Tr[**F**], which usually contains all the necessary information. It is then sufficient to compute the diagonal elements of $\boldsymbol{\mathcal{F}}$, since the two matrices, $\boldsymbol{\mathcal{F}}$ and $\mathbf{F}$, have the same trace. Using

$$\partial_{k_\alpha} u_{j\mathbf{k}} = \lim_{\Delta k_\alpha \to 0} \frac{u_{j,\mathbf{k}+\Delta k_\alpha} - u_{j\mathbf{k}}}{\Delta k_\alpha}. \tag{3.21}$$

we can write (substituting Δk_α with Δ_α for brevity)

$$\langle u_{n\mathbf{k}} | \partial_{k_\alpha} u_{n\mathbf{k}} \rangle = \lim_{\Delta_\alpha \to 0} \frac{\langle u_{n\mathbf{k}} | u_{n,\mathbf{k}+\Delta_\alpha} \rangle - 1}{\Delta_\alpha}. \tag{3.22}$$

Another finite difference expression results in

$$\partial_{k_\beta} \langle u_{n\mathbf{k}} | \partial_{k_\alpha} u_{n\mathbf{k}} \rangle = \lim_{\Delta_\beta \to 0} \lim_{\Delta_\alpha \to 0} \frac{\langle u_{n,\mathbf{k}+\Delta_\beta} | u_{n,\mathbf{k}+\Delta_\alpha+\Delta_\beta} \rangle - \langle u_{n\mathbf{k}} | u_{n,\mathbf{k}+\Delta_\alpha} \rangle}{\Delta_\alpha \Delta_\beta} \tag{3.23}$$

and for the curvature we get

$$\begin{aligned} \mathcal{F}_{nn,\alpha\beta}(\mathbf{k}) = i \lim_{\Delta_\beta \to 0} \lim_{\Delta_\alpha \to 0} \quad & \left[\frac{\langle u_{n,\mathbf{k}+\Delta_\alpha} | u_{n,\mathbf{k}+\Delta_\beta+\Delta_\alpha} \rangle - \langle u_{n\mathbf{k}} | u_{n,\mathbf{k}+\Delta_\beta} \rangle}{\Delta_\alpha \Delta_\beta} - \right. \\ & \left. - \frac{\langle u_{n,\mathbf{k}+\Delta_\beta} | u_{n,\mathbf{k}+\Delta_\alpha+\Delta_\beta} \rangle - \langle u_{n\mathbf{k}} | u_{n,\mathbf{k}+\Delta_\beta} \rangle}{\Delta_\alpha \Delta_\beta} \right]. \end{aligned}$$

Due to the fact that $\mathcal{F}_{nn,\alpha\beta}$ is purely real, the above expression can be rewritten

$$\begin{aligned} \mathcal{F}_{nn,\alpha\beta}(\mathbf{k}) = i \lim_{\Delta_\beta \to 0} \lim_{\Delta_\alpha \to 0} \frac{-1}{\Delta_\alpha \Delta_\beta} \operatorname{Im} \log & \left[\langle u_{n\mathbf{k}} | u_{n,\mathbf{k}+\Delta_\alpha} \rangle \langle u_{n,\mathbf{k}+\Delta_\alpha} | u_{n,\mathbf{k}+\Delta_\alpha+\Delta_\beta} \rangle \times \right. \\ & \left. \times \langle u_{n,\mathbf{k}+\Delta_\alpha+\Delta_\beta} | u_{n,\mathbf{k}+\Delta_\beta} \rangle \langle u_{n,\mathbf{k}+\Delta_\beta} | u_{n\mathbf{k}} \rangle \right] \end{aligned} \tag{3.24}$$

This expression is explicitly gauge-independent and is particularly simple for calculations on a discrete $\mathbf{k}$-mesh. The Chern number of Eq. (3.18) is obtained by a numerical integration of the above curvature and summing over all occupied bands when necessary.

3.3 Chern insulators

The first example of a topologically non-trivial insulator came in the context of the integer quantum Hall effect (IQH). The ordinary IQH is observed at low temperatures at in (effectively) two-dimensional semiconductor interfaces with high carrier mobility, subject to an in-plane electric field and a transverse magnetic field [65, 66]. The energy of a two-dimensional electron gas subject to a uniform transverse magnetic field is quantized into Landau levels [67]. Once the Fermi level of the IQH system is put in the gap between Landau levels (i.e., the occupied Landau levels are separated from the unoccupied ones by an energy gap) by either changing carrier concentration or the strength of magnetic field, the system exhibits a quantized transverse conductivity $\sigma_{xy} = Ce^2/h$, where C is an integer. This quantization was understood to have a topological origin. By explicit calculation of the Hall conductivity, Thouless and co-workers [12] were able to show that C is the Chern number analogous to that of Eq. (3.18), but with occupied Landau levels replacing the isolated Bloch bands and with integration over the magnetic BZ.[2]

There is an obvious analogy between an IQH system and an insulator due to the similarity between the filled Landau levels and the occupied states in an insulator. Electrons in the bulk of an IQH sample are localized just like in insulators [69], but for a finite IQH sample there exist states that are localized in one direction to be at the edge of the sample, but are extended along the edge [23]. These edge states are responsible for the quantized transport in the IQH regime, and the difference in the number of such states propagating in one direction along the edge and those propagating in the opposite direction is equal to the Chern number [66].

[2] Since in the general electromagnetic potential does not have the periodicity of the lattice, $\mathbf{A}(\mathbf{r}) \neq \mathbf{A}(\mathbf{r}+\mathbf{R})$, a unit cell has to be enlarged in the presence of an external magnetic field $\mathbf{B}$ in order to accommodate an integer number of the flux quanta eB/h through it. Hence, the magnetic BZ is reduced compared to the original one [68].

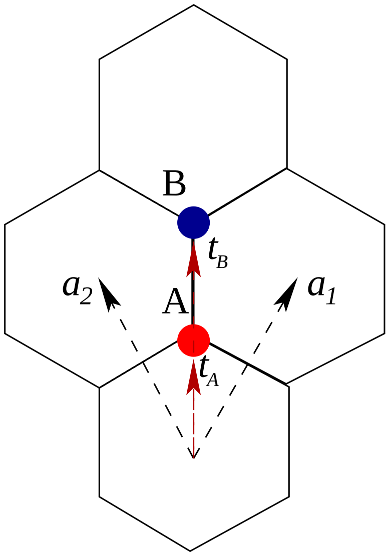

Figure 3.1: Honeycomb lattice for Haldene and Kane-Mele models.

The apparent analogy between an IQH system and an insulator hints at the possibility of realizing such edge states, and hence the IQH effect, in a band insulator in the absence of any external magnetic field. Being different from the usual IQH systems, these hypothetical insulators got the name of "quantum anomalous Hall insulators" (QAH) or "Chern insulators." Although at the time of writing this effect has not been observed experimentally in any material, it is easy to construct tight-binding models of such insulators. The first such model was proposed by Haldane [19], who considered a honeycomb lattice (see Fig. 3.1) with a single spinless electron per two sites (A and B) of the unit cell, described by the Hamiltonian

$$\hat{H}_{\mathrm{H}} = \lambda_{\mathrm{v}} \sum_{i} \xi_i \hat{c}_i^\dagger \hat{c}_i + t \sum_{<ij>} \hat{c}_i^\dagger \hat{c}_j + \lambda_{\mathrm{SO}} \sum_{\ll ij \gg} e^{i\alpha_{ij}} \xi_i \hat{c}_i^\dagger \hat{c}_j \tag{3.25}$$

with $\xi_i = 1$ on A sites and $\xi_i = -1$ on B sites, and symbols $<>$ and $\ll\gg$ referring to summations over first and second neighbors correspondingly. Here a macroscopic magnetic field is replaced by microscopic effective interactions.

These interactions are mimicked by the complex hopping term $\lambda_{\mathrm{SO}}e^{i\alpha_{ij}}$, having the effect of some microscopic magnetic field. The phase α_{ij} is designed to produce a non-zero magnetic flux through parts of the the unit cell, but the net flux through the unit cell is zero. It is chosen to be $\alpha_{ij} = \alpha$ if the hopping is in the clockwise direction (following the sides of a hexagon) and $\alpha_{ij} = -\alpha$ otherwise, which guarantees hermiticity. With this phase choice $\mathcal{T}$ symmetry of the system is broken, as is the case in the presence of an external magnetic field. The first term of the above Hamiltonian, staggered by the introduction of ξ_i, makes A and B sublattices inequivalent, thus breaking inversion symmetry.

The model is solved in the tight-binding basis of Sec. 2.3

$$\chi_{\mathbf{k}\ell} = \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}}\phi(\mathbf{r}-\mathbf{R}-\mathbf{t}_\ell), \tag{3.26}$$

where $\mathbf{t}_\ell$ gives the location the $\ell = \{A, B\}$ site in the home unit cell. Depending on the choice of parameters, the resultant $H(\mathbf{k})$ realizes three distinct insulating phases separated by gap closures. These phases correspond to three different values of the Chern number, 0 and ±1, obtained with the formula of Eq. (3.18) applied to the single occupied band of the model. We see that indeed, Berry curvature plays the role of a **B**-field in reciprocal space, in the sense that here singularities in $\mathcal{F}(\mathbf{k})$ give rise to the quantization of the Hall conductivity just as the **B**-field does in the IQH effect.

As in the case for the usual IQH effect, an edge of a Chern insulator hosts edge states in accord with the value of the Chern number. This is illustrated schematically in Fig. 3.2 for the case of a semi-infinite Haldane model (Fig. 3.3), where half of the two-dimensional space is the Chern insulator, while the other half is vacuum, the boundary between them being the edge. In the direction parallel to the edge the crystal is still periodic, so a the vector $k = k_{\parallel}$ is a good quantum number. One can look at the band structure (called the projected band

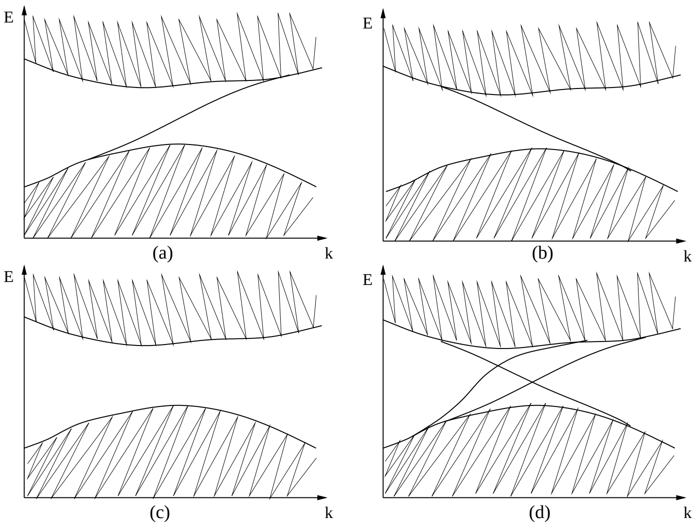

Figure 3.2: Schematic edge spectrum of the semi-infinite Chern insulator: (a) $C = 1$, (b) $C = -1$, (c) $C = 0$, (d) $C = 1$. The continuum of bulk bands is shaded.

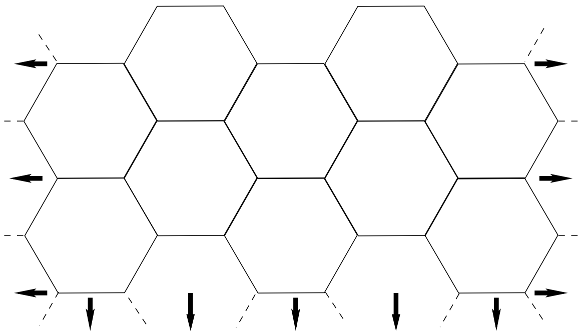

Figure 3.3: Edge of a semi-infinite sample. Arrows indicate continuation to infinity.

structure) of the system along this boundary. There will be infinite number of bulk bands and possibly some edge states present in the band structure. In the bulk insulating regime, only the edge states that can cross the energy gap.One can see exactly one edge state, propagating in different directions for the case of $C = \pm 1$, in accord with our previous discussion about the correspondence between the bulk topological invariant and the number of edge states in the IQH.

It should be noted that the band structure shown in Fig. 3.2 (d) also corresponds to $C = 1$. As mentioned above, it is the difference between the "right" and "left" moving edge states that corresponds to the Chern number. In the spirit of topology, this means that the band structures shown in Fig. 3.2 (a) and (d) are in the same topological class. This means that, in principle, these two structures can be adiabatically transformed into each other by, for example, tuning the parameters of the Hamiltonian or gradually adding some extra terms to it. On the contrary, two insulating phases with distinct topological invariants C cannot be adiabatically connected to one another without closing the bulk band gap.[3]

To conclude, we see that QAH phases are classified according to the value of Chern number which characterizes the occupied space of the insulator. Since this number can be any integer, there is an infinite number of distinct classes of QAH insulators. For this reason this classification is often referred to as the $\mathbb{Z}$-classification.

3.4 $\mathbb{Z}_2$ insulators

A series of theoretical developments starting in 2005, showing that non-magnetic insulators admit a topological $\mathbb{Z}_2$ classification in two dimensions (2D) [1, 24] and

[3] Strictly speaking, this argument requires the symmetry class of the system in the sense of Ref. [20] be preserved in the process of adiabatic change. We also limit ourselves to the case of a clean system (no disorder) and do not consider electron-electron interactions. However, one can argue that if the disorder or interaction strength is small compared to the band gap, the above classification survives.

then in three dimensions (3D) [28–30], has sparked enormous interest, especially after numerous realizations of such systems were confirmed both theoretically [25, 31, 70–75] and experimentally [26, 76–80].

The $\mathbb{Z}_2$ classification divides $\mathcal{T}$-invariant band insulators into two classes: ordinary ($\mathbb{Z}_2$-even) insulators that can be adiabatically converted to the vacuum (or to each other) without a bulk gap closure, and "topological" ($\mathbb{Z}_2$-odd) ones that cannot be so connected (although they can be adiabatically connected to each other).[4] Even and odd phases are separated by a topological phase transition, and the bulk gap has to vanish at the transition point, at least in a non-interacting system [20, 21]. The $\mathbb{Z}_2$-odd states are characterized by the presence of an odd number of Kramers pairs of counterpropagating edge states in 2D, or by an odd number of Fermi loops enclosing certain high-symmetry points of the surface band structure in 3D. In this subsection we review the theory behind the $\mathbb{Z}_2$-insulators.

3.4.1 Band structure in the presence of $\mathcal{T}$ symmetry

In the presence of $\mathcal{T}$ symmetry the energy bands have to come in Kramers pairs.[5] This means that for any band n there exists some other band m, such that

$$\epsilon_{n\mathbf{k}} = \epsilon_{m-\mathbf{k}} \tag{3.27}$$

and Hamiltonian eigenfunctions (which can be spinors) that correspond to a Kramers pair of energy bands n and m are related by the $\mathcal{T}$ operator

$$\theta|\psi_{n,-\mathbf{k}}\rangle = i\hat{s}_y\hat{K}|\psi_{m,\mathbf{k}}\rangle \tag{3.28}$$

[4] Here again the adiabatic connection is assumed to preserve the symmetry class in the sense of Ref. [20]. For example, it is possible to connect the $\mathbb{Z}_2$-odd and the $\mathbb{Z}_2$-even phases without closing the band gap by breaking $\mathcal{T}$ symmetry of the Hamiltonian in the process of adiabatic connection. However, breaking $\mathcal{T}$ symmetry drives the system out of the original symmetry class.

[5] This statement is known as the Kramers theorem [2].

where $\hat{K}$ is the complex conjugation and $\hat{s}_y$ is the y-component of the spin operator.

It follows from this relation that there exist points in the BZ where the energy spectrum is at least doubly degenerate, degeneracy being protected by $\mathcal{T}$ symmetry. Indeed, there are points $\mathbf{k}^*$ such that $-\mathbf{k}^* = \mathbf{k}^* + \mathbf{G}$, and due to periodicity of BZ, $\epsilon_{n\mathbf{k}^*} = \epsilon_{n-\mathbf{k}^*} = \epsilon_{m-\mathbf{k}^*}$. Since $\mathcal{T}$ maps $\mathbf{k}$ onto $-\mathbf{k}$, such points $\mathbf{k}^*$ are called $\mathcal{T}$-invariant points. There are four distinct $\mathcal{T}$-invariant points in a 2D BZ ($\mathbf{k}^* = 0$; $\mathbf{k}^* = \mathbf{G}_1/2$; $\mathbf{k}^* = \mathbf{G}_2/2$ and $\mathbf{k}^* = (\mathbf{G}_1 + \mathbf{G}_2)/2$), while in 3D there are eight such points ($\mathbf{k}^* = 0$; $\mathbf{k}^* = \mathbf{G}_i/2$; $\mathbf{k}^* = (\mathbf{G}_i + \mathbf{G}_j)/2$, where $i, j = 1, 2, 3$ and $\mathbf{k}^* = (\mathbf{G}_1 + \mathbf{G}_2 + \mathbf{G}_3)/2$). Another consequence of the Kramers relations is that the occupied space of an insulator consists of an even number of bands, since the occupied states come in pairs.

3.4.2 2D: Quantum spin Hall phase

Let us now include spin degrees of freedom into the Haldane model (3.25). In the case when $\mathcal{T}$ symmetry remains broken, sinfull insulating phases will still be classified according to the Chern number of the occupied space. However, if we restore $\mathcal{T}$ symmetry, total Chern number is guaranteed to be zero, but a new phase can arise. For a moment, consider a simplified model where $\hat{s}_z$ is a conserved quantity. $\mathcal{T}$-symmetric sinfull generalization of Haldane model takes the form of a 4×4 block-diagonal matrix Hamiltonian

$$H_{\mathrm{SH}} = \begin{pmatrix} H_{\mathrm{H}}(\mathbf{k}) & 0 \\ 0 & H^*_{\mathrm{H}}(-\mathbf{k}) \end{pmatrix}, \tag{3.29}$$

where diagonal blocks are the 2×2 Hamiltonians of the Haldane model (3.25). Obviously, the occupied space of such a system will consist of two bands of opposite spin, each of them having a definite Chern number, inherited from the

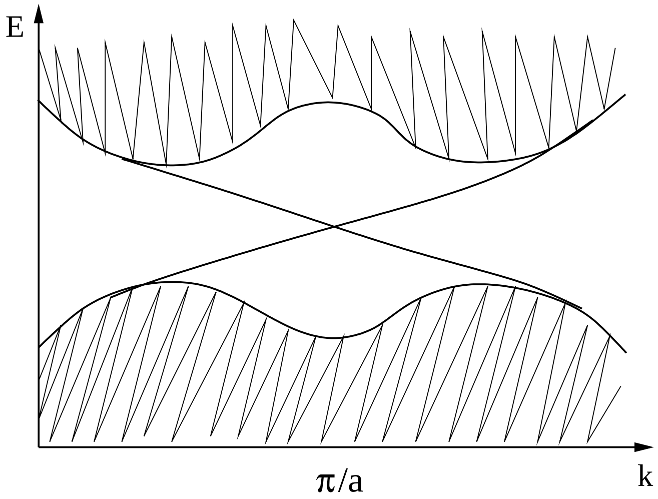

Figure 3.4: Band structure for a hypothetical semi-infinite sample of the quantum spin Hall insulator. Shadowed regions correspond to the continuum of bulk states.

Haldane model.

Note that in principle, it does not make sense to talk about Chern numbers of separate bands in the presence of degeneracy. However, in this case the spin quantum number allows us to distinguish two bands, and define a Chern number for each of them, using the formula (3.18) with $\mathcal{F}$ computed separately for spin up and spin down states. The Hall conductivity is odd under $\mathcal{T}$, so that the total Chern number of all the occupied bands in a $\mathcal{T}$-symmetric insulator has to be zero. Therefore, the Chern number of the spin-up state is minus that of the spin-down state

$$C_{\uparrow} = -C_{\downarrow}. \tag{3.30}$$

It is interesting to look at what happens at the edge of such a system. In the Sec. 3.3 it was argued that a semi-infinite sample of a Chern insulator is characterized by the presence of current-carrying edge states. For the $\mathcal{T}$-symmetric sinfull case, the typical spectrum of the semi-infinite model is shown in Fig. 3.4. There are two edge states that have opposite spins and propagate in different directions, as a consequence of Eq. (3.30). This means that there is no charge transport along the edge of the sample, but there is a spin current instead. Moreover, due to the conservation of spin, this spin current is quantized, and the corresponding spin Hall conductivity is [24]

$$\sigma_{xy}^{\mathrm{sp}} = \frac{(C_\uparrow - C_\downarrow)}{2}\frac{e}{2\pi} = C_{\mathrm{sp}}\frac{e}{2\pi}, \tag{3.31}$$

where we defined a spin Chern [81] number C_{sp}. We see that, in principle, a $\mathcal{T}$-symmetric combination of Chern insulators results in a quantized spin Hall effect [24]. This is very intriguing, given the quest for controlled spin transport in spintronics [82].

At first sight it might look as though $\mathcal{T}$-symmetric phases are classified by the value of C_{sp} in complete analogy with the IQH. However, it is easy to see that such a viewpoint leads to certain problems. Unlike charge, the spin projection s_z is not generally conserved. For example, the spin-orbit coupling usually violates spin conservation. Thus, the distinction between the phase with C_{sp} and $-C_{\mathrm{sp}}$ is lost once a spin-breaking term is appears in the Hamiltonian. In that case it is possible to change the sign of the spin Hall conductivity without closing the bulk gap [27]. This immediately tells us that, from a topological perspective, these two phases should be in the same class. Finally, the most important observation is that the spin transport at the edge is robust towards small perturbations only in the case of $C_\uparrow = -C_\downarrow$ being odd.

To see this, notice that in the case of $C_\uparrow = -C_\downarrow = 1$ illustrated in Fig. 3.4, $\mathcal{T}$ symmetry guarantees the crossing of the two surface states. At $k = \pi/a$ Kramers theorem forces the spectrum to be doubly degenerate, and thus no adiabatic transformation that preserves $\mathcal{T}$ can open up a gap in the surface spectrum. Quite the contrary, consider $C_\uparrow = -C_\downarrow = 2$. Fig. 3.5 illustrates the possible scenario for gapping the surface states in this case. Note that the condition of doubly degenerate energy bands at the $\mathcal{T}$-symmetric momenta is still satisfied.

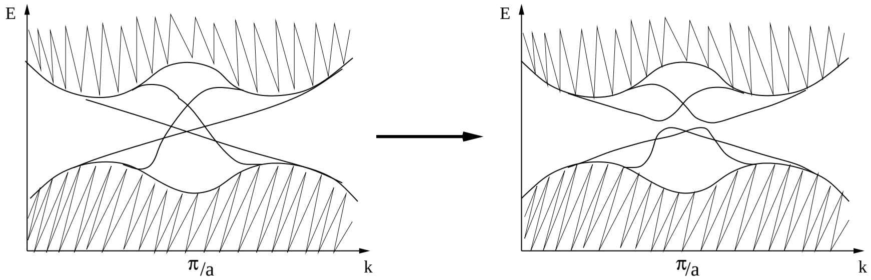

Figure 3.5: Gapping the edge spectrum by adiabatic changes in the Hamiltonian.

Thus, the Hamiltonian can be adiabatically tuned to create avoided crossings in the surface spectrum, and furthermore, one can tune the Hamiltonian further and push the split surface states into the bulk region, thus making an adiabatic connection to an ordinary insulator. By the same token, the case of any odd $C_\uparrow$ can be adiabatically connected to the case of $C_\uparrow = 1$, while for any even value of $C_\uparrow$ the system is adiabatically equivalent to an ordinary insulator with no surface states. Thus, we conclude that there are only two distinct phases (odd vs even) that cannot be smoothly deformed into one another. For this reason this classification is called a $\mathbb{Z}_2$-classification [1], since $\mathbb{Z}_2$ is a group that consists of two elements only. We now show that this classification survives the addition of spin-mixing terms.

3.4.3 Kane-Mele model

In their remarkable paper introducing a $\mathbb{Z}_2$ topological classification to distinguish a QSH ($\mathbb{Z}_2$-odd) insulator from an ordinary ($\mathbb{Z}_2$-even) insulator, Kane and Mele (KM) [1] also introduced a model tight-binding Hamiltonian that describes a 2D $\mathbb{Z}_2$-odd insulator in some of its parameter space. In this section we will describe some of the properties of the model suggested therein.

The KM model is a tight-binding model on a honeycomb lattice with one spinor orbital per site. The primitive hexagonal lattice vectors are $\mathbf{a}_{1,2} = a/2(\sqrt{3}\hat{\mathbf{y}}\pm$

$\hat{\mathbf{x}}$) and sites A and B are located at $\mathbf{t}_A = a\hat{y}/\sqrt{3}$ and $\mathbf{t}_B = 2a\hat{y}/\sqrt{3}$ respectively (see Fig. 3.1). The KM Hamiltonian is

$$H = t\sum_{<ij>} c_i^\dagger c_j + i\lambda_{\rm SO}\sum_{\ll ij\gg} \nu_{ij} c_i^\dagger s^z c_j + i\lambda_{\rm R}\sum_{<ij>} c_i^\dagger (\mathbf{s}\times\hat{\mathbf{d}}_{ij})_z c_j + \lambda_{\rm v}\sum_i \xi_i c_i^\dagger c_i, \quad (3.32)$$

where the spin indices have been suppressed on the raising and lowering operators, and t is the nearest-neighbor hopping amplitude. In the second term, $\lambda_{\rm SO}$ is the strength of the spin-orbit interaction acting between second neighbors, with $\nu_{ij} = (2/\sqrt{3})[\hat{\mathbf{d}}_1 \times \hat{\mathbf{d}}_2] = \pm 1$ depending on the relative orientation of the first-neighbor bond vectors $\hat{\mathbf{d}}_1$ and $\hat{\mathbf{d}}_2$ encountered by an electron hopping from site j to site i, and s^z is the z Pauli spin matrix. Next, $\lambda_{\rm R}$ describes the Rashba interaction [83] that couples differently oriented first-neighbor spins, with $\mathbf{s}$ being the vector of Pauli matrices. Finally, $\lambda_{\rm v}$ is the strength of the staggered on-site potential, for which ξ_i is $+1$ and -1 on A and B sites respectively. Note that the symmetry of the problem is lowered significantly compared to an ideal honeycomb lattice, since the on-site staggered potential makes the A and B sites inequivalent, while the Rashba term breaks s^z conservation.

To proceed, we choose the tight-binding basis wavefunctions to be

$$\chi_{j\sigma\mathbf{k}}(\mathbf{r}) = (1/\sqrt{N})\sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}}\phi_\sigma(\mathbf{r}-\mathbf{R}-\mathbf{t}_j), \quad (3.33)$$

where σ is a spin index, $j = \{A, B\}$ denotes the atom type, $\mathbf{t}_j$ is a vector that specifies the position of the atom in the unit cell, and $\mathbf{R}$ is a lattice vector built from the primitive lattice vectors $\mathbf{a}_1$ and $\mathbf{a}_2$. This allows the Hamiltonian to be written as a 4×4 matrix $H_{j\sigma,j'\sigma'}(\mathbf{k}) = \langle\chi_{j\sigma\mathbf{k}}|H|\chi_{j'\sigma'\mathbf{k}}\rangle$, which can be cast in terms of five Dirac matrices Γ^α and their ten commutators $\Gamma^{\alpha\beta} = [\Gamma^\alpha, \Gamma^\beta]/(2i)$ as

$$H(\mathbf{k}) = \sum_{\alpha=1}^{5} d_\alpha(\mathbf{k})\,\Gamma^\alpha + \sum_{\alpha<\beta=1}^{5} d_{\alpha\beta}(\mathbf{k})\,\Gamma^{\alpha\beta} \quad (3.34)$$

d_1	$t(1+2\cos x\cos y)$	d_{12}	$-2t\cos x\sin y$
d_2	$\lambda_{\rm v}$	d_{15}	$2\lambda_{\rm SO}(\sin 2x-2\sin x\cos y)$
d_3	$\lambda_{\rm R}(1-\cos x\cos y)$	d_{23}	$-\lambda_{\rm R}\cos x\sin y$
d_4	$-\sqrt{3}\lambda_{\rm R}\sin x\sin y$	d_{24}	$\sqrt{3}\lambda_{\rm R}\sin x\cos y$

Table 3.1: Nonzero coefficients appearing in Eq. (3.34), using the notation $x = k_xa/2$ and $y = \sqrt{3}k_ya/2$ (see also Fig. 3.6). Adopted from Ref. [1].

where the Dirac matrices are chosen to be $\Gamma^{1,2,3,4,5} = (I\otimes\sigma^x, I\otimes\sigma^z, s^x\otimes\sigma^y, s^y\otimes\sigma^y, s^z\otimes\sigma^y)$ with the Pauli matrices σ^ℓ and s^ℓ acting in sublattice and spin space respectively. The dependence of the d_α and $d_{\alpha\beta}$ coefficients on wavevector is detailed in Table 3.1 using the notation $x = k_xa/2$ and $y = \sqrt{3}k_ya/2$, with the relationship of these variables to the BZ being sketched in Fig. 3.6.

Since $\hat\theta\Gamma^\alpha\hat\theta^{-1} = \Gamma^\alpha$ and $\hat\theta\Gamma^{\alpha\beta}\hat\theta^{-1} = -\Gamma^{\alpha\beta}$ while $d_\alpha(\mathbf{k}) = d_\alpha(-\mathbf{k})$ and $d_{\alpha\beta}(\mathbf{k}) = -d_{\alpha\beta}(-\mathbf{k})$, the Hamiltonian (3.32) is time-reversal invariant, i.e., $\hat\theta H(\mathbf{k})\hat\theta^{-1} = H(-\mathbf{k})$. This means that its insulating phases are classified by the $\mathbb{Z}_2$ index. Indeed, in a certain parameter range a semi-infinite slab of the model exhibits an edge spectrum like that schematically shown in Fig. 3.4, while for other values of parameters no edge states cross the bulk gap. On the boundary of these phases the system becomes metallic, signaling a topological phase transition. Thus, we conclude that the this model has a $\mathbb{Z}_2$ classification. This is confirmed by a direct calculation of a corresponding topological invariant (see below). Since s_z is not conserved anymore, the spin current is also not quantized, and we are dealing with a quantum (not quantized) spin Hall phase (QSH).

For the present purposes we assume $\lambda_{\rm SO} > 0$ without loss of generality. We also fix $\lambda_{\rm v} > 0$. For this case, the transition between $\mathbb{Z}_2$-odd and $\mathbb{Z}_2$-even phases is accompanied by a gap closure at the K and K' points (the zone-boundary points of three-fold symmetry) in the BZ. The energy is independent of t at these points, and $\lambda_{\rm SO}$ can be used as the energy scale. The energy gap is then given by

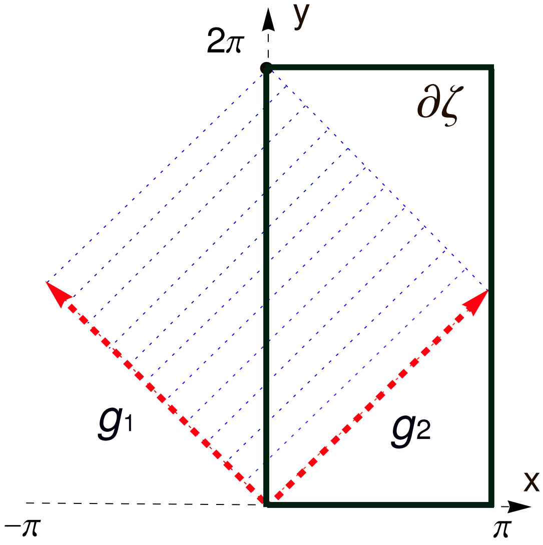

Figure 3.6: Brillouin zone sketched using coordinates $x = k_x a/2$ and $y = \sqrt{3}k_y a/2$. Primitive reciprocal lattice vectors $\mathbf{G}_1 = (2\pi/a)(1, 1/\sqrt{3})$ and $\mathbf{G}_2 = (2\pi/a)(-1, 1/\sqrt{3})$ correspond to $\mathbf{g}_1 = (\pi, \pi)$ and $\mathbf{g}_2 = (-\pi, \pi)$ respectively. The black rectangle marks the boundary $\partial\zeta$ of the zone used for polarization calculations in Sec. 6.3.

$|6\sqrt{3} - \lambda_{\rm v}/\lambda_{\rm SO} - \sqrt{(\lambda_{\rm v}/\lambda_{\rm SO})^2 + 9(\lambda_{\rm R}/\lambda_{\rm SO})^2}|$, leading to the phase diagram shown in Fig. 3.7.

Note that when $\lambda_{\rm R} = 0$ the model reduces to two independent copies of the Haldane model [19]; the $\mathbb{Z}_2$ invariant is odd when the Chern numbers are odd, and even otherwise [81], in accord with the discussion in Sec. 3.3.

In what follows we use t as the energy scale and fix the values of the other parameters to be $\lambda_{\rm SO}/t = 0.6$ and $\lambda_{\rm R}/t = 0.5$. Varying the third parameter $\lambda_{\rm v}/t$ allows us to switch from the $\mathbb{Z}_2$-even to the $\mathbb{Z}_2$-odd phase. The phase transition occurs at $|\lambda_{\rm v}|/t \simeq 2.93$, with the system in the $\mathbb{Z}_2$-odd phase for $-2.93 < \lambda_{\rm v}/t < 2.93$. As discussed above, the energy gap closes at the phase transition, and remains open in both the $\mathbb{Z}_2$-odd and $\mathbb{Z}_2$-even phases. When referring to a particular phase of the KM model, in numerical calculations we use $\lambda_{\rm v}/t = 1$ for the $\mathbb{Z}_2$-odd phase and $\lambda_{\rm v}/t = 5$ in the $\mathbb{Z}_2$-even (normal insulator) phase.

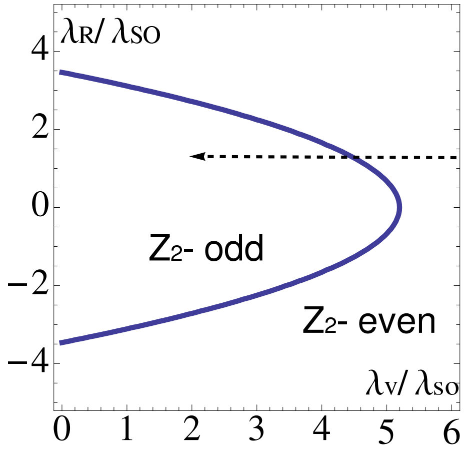

Figure 3.7: Phase diagram of the Kane-Mele model for $\lambda_{\mathrm{v}}/\lambda_{\mathrm{SO}} > 0$. Arrow illustrates a path crossing the phase boundary by varying λ_{v} while keeping other parameters fixed.

3.4.4 Computing the $\mathbb{Z}_2$ invariant

Here we briefly review some of the equivalent ways of determining the $\mathbb{Z}_2$ invariant in 2D insulators.

In the work of Ref. [24] the definition of the $\mathbb{Z}_2$ invariant was given in terms of a function $P(\mathbf{k})$ defined as

$$P(\mathbf{k}) = \mathrm{Pf}[\langle u_i(\mathbf{k})|\hat{\theta}|u_j(\mathbf{k})\rangle], \tag{3.35}$$

i.e., the Pfaffian of a certain $\mathbf{k}$-dependent antisymmetric $N \times N$ matrix, where N is the number of occupied bands. Here $|u_j(\mathbf{k})\rangle = e^{-i\mathbf{k}\cdot\mathbf{r}}|\psi_j(\mathbf{k})\rangle$ is the periodic part of the Bloch function of the j'th occupied band and $\hat{\theta} = is^y\hat{C}$ is the time-reversal operator ($\hat{C}$ is complex conjugation and s^y is the second Pauli matrix). If the zeros of $P(\mathbf{k})$ are discrete, then the $\mathbb{Z}_2$ invariant is odd if the number of zeros of the Pfaffian within one half of the Brillouin zone (BZ) (see Fig. 3.8) is odd, and even otherwise. If the zeros of the Pfaffian occur along lines in the BZ, then the $\mathbb{Z}_2$ invariant depends similarly on whether half the number of sign changes of

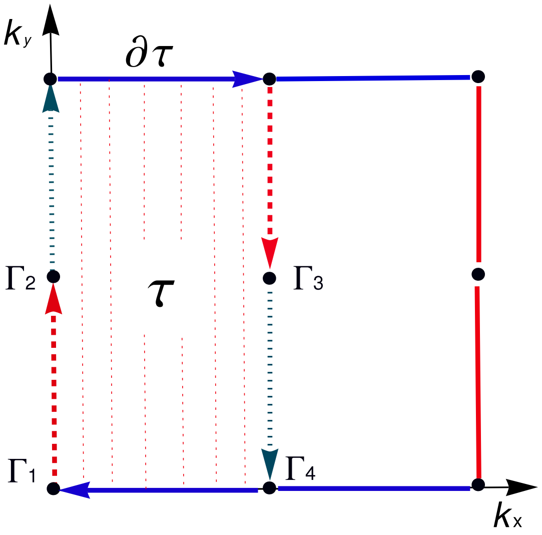

Figure 3.8: Sketch of the Brillouin zone. The Berry curvature of Eq. (3.38) is calculated in the interior of the half zone τ (dashed region), while the Berry connection is evaluated along its boundary $\partial\tau$ (arrows indicate direction of integration). Time-reversal–invariant points Γ_i are shown.

$P(\mathbf{k})$ along the boundary of the half BZ is odd or even. Using $\Delta = 0$ and 1 to represent evenness and oddness respectively, the $\mathbb{Z}_2$ invariant can equivalently be determined as [1]

$$\Delta = \frac{1}{2i\pi}\oint_{\partial\tau} d\mathbf{k}\cdot\nabla_{\mathbf{k}}\log[P(\mathbf{k}+i\delta)] \mod 2, \tag{3.36}$$

where the loop integral runs along the boundary $\partial\tau$ of the half BZ, and the δ term is included for convergence.

Another approach to the problem of defining Δ results from considerations of “time-reversal polarization” [27] (see Chapter 5 for details). This approach leads to the formula

$$(-1)^{\Delta} = \prod_{i=1}^{4}\frac{\sqrt{\det[w(\mathbf{\Gamma}_i)]}}{\mathrm{Pf}[w(\mathbf{\Gamma}_i)]}, \tag{3.37}$$

where $w_{mn}(\mathbf{k}) = \langle u_m(-\mathbf{k})|\hat{\theta}|u_n(\mathbf{k})\rangle$ and $\mathbf{\Gamma}_i$ are the four $\mathcal{T}$-invariant points of the BZ. The matrix w_{mn} is not the same as that in Eq. (3.35). Note that $w(-\mathbf{k}^*) =$

$-w^T(\mathbf{k}^*)$, so that the Pfaffian in (3.37) is well defined.

The definition in Eq. (3.37) appears to require a knowledge of the occupied wavefunctions at only four points in the BZ, unlike Eq. (3.36), for which the wavefunctions must be known at all points along the boundary of the half BZ. However, Eq. (3.37) is usually not suitable for numerical implementation in practice, since the sign of the Pfaffian at any one of the four points can be flipped by a relabeling of the Kramers-degenerate states at that point. To be more explicit, there is a "gauge freedom" in the choice of states $|u_m(\mathbf{k})\rangle$, corresponding to a $\mathbf{k}$-dependent $N \times N$ unitary rotation among the occupied states. Eq. (3.37) is only meaningful when a globally smooth gauge choice enforces a relation between the labels at the four special $\mathbf{k}$-points [27]. This problem may be avoided in the presence of some additional symmetry that can be used to establish the labels of the bands at these points. For example, in Ref. [31] it is shown how the presence of inversion symmetry allows for a simplified calculation of Δ from Eq. (3.37). In Chapter 7 we show how Δ can be computed using the formula (3.37) in the general case.

In the absence of inversion symmetry, one can use yet another definition of the $\mathbb{Z}_2$ index taking the form [27]

$$\Delta = \frac{1}{2\pi}\left[\oint_{\partial\tau} \mathcal{A} d\ell - \int_{\tau} \mathcal{F} d\tau\right] \quad \text{mod } 2, \tag{3.38}$$

where $\mathcal{A} = i\sum_{n=1}^{\mathcal{N}} \langle u_n|\nabla_{\mathbf{k}}|u_n\rangle$ is the Berry connection of $\mathcal{N}$ occupied states and $\mathcal{F} = \nabla_{\mathbf{k}} \times \mathcal{A}$ is the corresponding Berry curvature [6]. Of course, if $\mathcal{A}$ and $\mathcal{F}$ are both constructed from a common gauge that is smooth over τ, the result would vanish by Stokes' theorem. Thus, Eq. (3.38) is only made meaningful by the additional specification [27] that the boundary integral of $\mathcal{A}$ must be calculated using a gauge that respects time-reversal symmetry, i.e.,

$$\begin{aligned} |u_{2n-1}(-\mathbf{k})\rangle &= \hat{\theta}|u_{2n}(\mathbf{k})\rangle, \\ |u_{2n}(-\mathbf{k})\rangle &= -\hat{\theta}|u_{2n-1}(\mathbf{k})\rangle. \end{aligned} \tag{3.39}$$

For the case of the nontrivial $\mathbb{Z}_2$ state, it turns out to be impossible to choose a gauge that satisfies both smoothness over τ and the constraint (3.39) over $\partial\tau$. In other words, Δ=1 signals the existence of the topological obstruction.

To see how this works more explicitly, the contributions to the integral of $\mathcal{A}$ over $\partial\tau$ are illustrated in Fig. 3.8. We choose a gauge that is periodic, $|u_j(\mathbf{k})\rangle = |u_j(\mathbf{k}+\mathbf{G})\rangle$, in addition to satisfying Eq. (3.39). The contributions of the top and bottom segments (solid blue arrows in Fig. 3.8) then cancel because they are connected by a reciprocal lattice vector $\mathbf{G}$. Thus, the gauge needs to be fixed only along the left and right boundaries (composed of red dashed and gray dotted arrows in Fig. 3.8), which are separated by a half reciprocal lattice vector. At each of the special points Γ_i, one state from each Kramers-degenerate pair is arbitrarily identified as $|u_{2n-1}(\Gamma_i)\rangle$, and the other is constructed via

$$|u_{2n}(\Gamma_i)\rangle = -\hat{\theta}|u_{2n-1}(\Gamma_i)\rangle. \tag{3.40}$$

Then we can make an arbitrary gauge choice along the remaining portions of the gray dotted arrows in Fig. 3.8 – e.g., accepting the output of some numerical diagonalization procedure. Finally, the gauge should be transferred to the dashed-arrow segments using Eq. (3.39), where $\mathbf{k}$ and $-\mathbf{k}$ belong to the dotted and dashed segments respectively.

Eq. (3.38) can now be evaluated using a uniform discretized $\mathbf{k}$-mesh covering the region τ, with the time-reversal constraint applied to the boundary $\partial\tau$ as described above. To do so, define the link matrices $M_{\mu,nm}(\mathbf{k}) = \langle u_n(\mathbf{k})|u_m(\mathbf{k}+\mathbf{s}_\mu)\rangle$

and the unimodular link variables $L_\mu(\mathbf{k}) = \det M_\mu / |\det M_\mu|$, where $\mathbf{k} \in \mathbb{K}$ and $\mathbf{s}_1$ ($\mathbf{s}_2$) is the step of the mesh in the direction of the reciprocal lattice vector $\mathbf{G}_1$ ($\mathbf{G}_2$). By defining $A_1(\mathbf{k}) = \log L_1(\mathbf{k})$ and

$$F(\mathbf{k}) = \log[L_1(\mathbf{k}) L_2(\mathbf{k} + \mathbf{s}_1) L_1^{-1}(\mathbf{k} + \mathbf{s}_2) L_2^{-1}(\mathbf{k})], \tag{3.41}$$

one can write the lattice definition of the $\mathbb{Z}_2$ invariant as

$$\Delta_L = \frac{1}{2i\pi} \left[\sum_{\mathbf{k} \in \partial\mathcal{T}} A_1(\mathbf{k}) - \sum_{\mathbf{k} \in \mathcal{T}} F(\mathbf{k}) \right] \mod 2. \tag{3.42}$$

For a sufficiently fine mesh there will be no ambiguity in the branch choice for the complex log in Eq. (3.41), since the argument of the log must approach unity as the mesh becomes dense. Moreover, a change in the branch choice determining one of the boundary links $A_s(\mathbf{k})$ has no effect (mod 2) on Eq. (3.41), since each $A_s(\mathbf{k})$ appears twice as a result of the gauge-fixing on the boundary. Thus, once the mesh is fine enough so that the branch choices in Eq. (3.41) are all unambiguous, Eq. (3.42) gives Δ exactly [84]. Although easy to implement in the TBA models, this method is not convenient for large scale first-principles calculations. In Chapter 5 we will develop another method for computing $\mathbb{Z}_2$ index in a non-centrosymmetric system that is better suited for *ab initio* applications.

3.4.5 3D: surface metal

A generalization of a 2D $\mathbb{Z}_2$ insulator to 3D is quite different from the 3D Chern insulator discussed above. For Chern insulators a topological invariant C_α is well defined at any 2D plane orthogonal to $\hat{k}_\alpha$, while a $\mathbb{Z}_2$ invariant is defined only for $\mathcal{T}$-invariant planes, e.g. for the values of $k_\alpha = k^*$ that are invariant under $\mathcal{T}$. Thus, the argument that the topological structure of a 2D cross section cannot change while adiabatically changing k_α is not applicable anymore, and different planes at different $\mathcal{T}$-symmetric momenta can have different $\mathbb{Z}_2$ indices in a 2D sense, making the classification more diverse than in the Chern insulator case.

A topological phase of a 3D $\mathcal{T}$-symmetric insulator is described by one strong topological index ν_0 and three weak indices ν_1, ν_2, and ν_3 [28–30]. These indices may be understood as follows. Again letting $\mathbf{k} = \sum_i k_i \mathbf{b}_i/2\pi$, there are eight $\mathcal{T}$-invariant points $\Gamma_{(n_1,n_2,n_3)}$, where $n_i = 0$ or 1 denotes $k_i = 0$ or π respectively. These eight points may be thought of as the vertices of a parallelepiped in reciprocal space whose six faces are labeled by n_1=0, n_2=0, n_3=0, n_1=1, n_2=1, and n_3=1. On any one of these six faces, the Hamiltonian $H(\mathbf{k})$, regarded as a function of two k variables, can be thought of as the Hamiltonian of a fictitious 2D $\mathcal{T}$-symmetric system, and the argument of the previous paragraph can thus be applied to each of these six faces separately. The three weak indices $\nu_{i=1,2,3}$ are defined to be the $\mathbb{Z}_2$ invariants associated with the three surfaces n_1=1, n_2=1, and n_3=1 [28]. These weak indices obviously depend on the choice of the reciprocal lattice vectors. The strong index ν_0 is the sum (mod 2) of the $\mathbb{Z}_2$ invariants of the n_j=0 and n_j=1 faces for any one of the j (implying some redundancy among the six indices); it is also a $\mathbb{Z}_2$ quantity, but is independent of the choice of reciprocal lattice vectors [28, 30].

Instead of the topologically protected 1D edge states in a 2D case, the boundary of a 3D material is a 2D surface. In $\mathbb{Z}_2$ topological insulators these surfaces are necessarily metallic. However, this metal is quite different from normal 2D metals [15]. First of all, it was argued that the surface states of a strong topological insulator (TI) are robust to weak non-magnetic disorder [85, 86], while ordinary 2D metals get localized in the presence of disorder [87, 88]. Moreover, recent results show that even for a weak TI the surface remains conducting in the presence of weak disorder [89, 90]. Moreover, the surface of a TI was predicted to exhibit a plethora of effects, not observed with ordinary metals/insulators. Examples include, but are not limited to, a quantized electromagnetic response [91, 92], fascinating optical effects [93, 94], realization of magnetic monopoles [95] and new possibilities for quantum computation [96].

Chapter 4

Wannier functions

We have seen that band theory usually operates in terms of Bloch states that are extended in real space. As usual in quantum mechanics, other representations for the wavefunctions are also possible, and the choice of a particular one to work in is a matter of convenience. For many problems in band theory, Bloch states provide a natural framework, being localized in reciprocal space. However, in certain cases it is much easier to work in a real-space representation, in which the wavefunctions are localized in real space. Insulators serve as a good example here. The electronic charge is localized in the insulating state [69, 97, 98], so it is natural to work with wavefunctions that are localized and can describe the the correct charge distribution. Although the convenience of such a representation is obvious, ways to construct it remained unclear for a long time.

A possible approach was introduced by Wannier [99] who considered Fourier transformed Bloch states, that now hold the name of Wannier functions (WFs),

$$|\mathbf{R}n\rangle = \frac{V_{\text{cell}}}{(2\pi)^2} \int_{BZ} d\mathbf{k}\, e^{-i\mathbf{k}\cdot\mathbf{R}} |\psi_{n\mathbf{k}}\rangle, \tag{4.1}$$

where V_{cell} is the unit cell volume and Bloch wavefunctions $\psi_{n\mathbf{k}}$ are assumed to be normalized within the unit cell. The label $\mathbf{R}$ gives the location of the unit cell to which WF belongs and n distinguishes different WFs. Although WFs themselves are not periodic functions of $\mathbf{r}$, they all have periodic images in the other unit cells, which can be obtained using the lattice translation operators $|\mathbf{R}n\rangle = \hat{T}_{\mathbf{R}}|\mathbf{0}n\rangle$ (see Fig. 4.1).

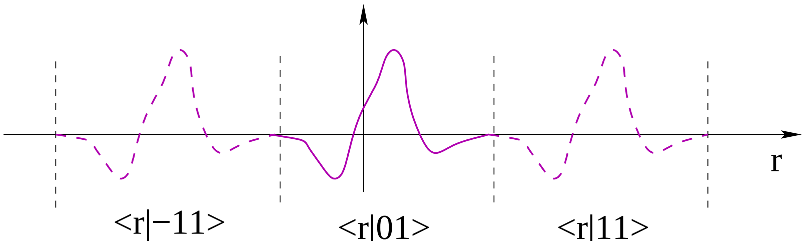

Figure 4.1: A sketch of amplitudes of Wannier functions in a 1D crystal. Solid curve refers to the home unit cell; dashed curves refer to periodic images in the neighboring cells.

The potential effectiveness of WFs in problems with localized electronic charge can be seen from the following consideration. Define the Wannier charge centers (WCC) to be the centers of mass of WFs:

$$\bar{\mathbf{r}}_n = \langle \mathbf{0}n|\hat{\mathbf{r}}|\mathbf{0}n\rangle. \tag{4.2}$$

If WFs are good candidates for a useful real-space description of solids, the location of the WCC should be close (at least in some sense) to the center of mass of the electronic density of a real material. Besides, the amplitude of a WF should fall off away from the WCC [100, 101], as is the case for the density of a localized charge distribution.

In principle, the WFs of Eq. (4.1) can be tuned to satisfy the above conditions, since as defined they are not unique. Indeed, as discussed in Sec. 3.1, any set of Bloch-like states $|\psi_{n\mathbf{k}}\rangle$ that span the occupied space of the problem can be used to construct WFs. In fact, it is generally necessary to apply a U($\mathcal{N}$) transformation of the form (3.5) to the Hamiltonian eigenstates in order that the resulting Bloch-like states (and their phases) are smooth functions of $\mathbf{k}$. However, having done so, there is still a large gauge freedom associated with the application of a subsequent $\mathcal{U}(\mathcal{N})$ gauge rotation that is smooth in $\mathbf{k}$. In general, the localization properties and the locations of the WCCs will be different for different choices of $|\psi_{n\mathbf{k}}\rangle$.

This ambiguity in the gauge choice can be removed by applying some criterion to the choice of the WFs. A sensible criterion is that of Ref. [32], which specifies maximal localization of the WFs in real space, as reviewed further.

4.1 Maximally localized Wannier functions

Taking into account the above mentioned properties that we want to see in WFs, it is natural to consider a choice of gauge that results in maximal localization of WFs. The problem of constructing maximally-localized WFs was studied by Marzari and Vanderbilt [32]. They considered the total quadratic spread

$$\Omega = \sum_{n=1}^{\mathcal{N}} [\langle \mathbf{0}n|\hat{r}^2|\mathbf{0}n\rangle - \langle \mathbf{0}n|\hat{\mathbf{r}}|\mathbf{0}n\rangle^2] \tag{4.3}$$

as a measure of the delocalization of WFs in real space, and developed methods for iteratively reducing the spread via a series of unitary transformations, Eq. (3.5), applied prior to WF construction. The spread functional was decomposed into two parts, $\Omega = \Omega_I + \tilde{\Omega}$, with

$$\Omega_I = \sum_{n=1}^{\mathcal{N}} \left[\langle \mathbf{0}n|\hat{r}^2|\mathbf{0}n\rangle - \sum_{m=1}^{\mathcal{N}} \sum_{\mathbf{R}} |\langle \mathbf{R}m|\hat{\mathbf{r}}|\mathbf{0}n\rangle|^2 \right] \tag{4.4}$$

being the gauge-invariant part and

$$\tilde{\Omega} = \sum_{n=1}^{\mathcal{N}} \sum_{\mathbf{R}m \neq \mathbf{0}n} |\langle \mathbf{R}m|\hat{\mathbf{r}}|\mathbf{0}n\rangle|^2 \tag{4.5}$$

the gauge-dependent part of the spread. Discretized **k**-space formulas for Eqs. (4.4) and (4.5) were also derived for the case that the BZ is represented by a uniform **k** mesh. The resulting expression for the gauge-invariant spread is, for example,

$$\Omega_I = \frac{1}{N} \sum_{\mathbf{k},\mathbf{b}} \omega_b \sum_{m,n=1}^{\mathcal{N}} \left(\delta_{mn} - |M_{mn}^{(\mathbf{k},\mathbf{k}+\mathbf{b})}|^2 \right), \tag{4.6}$$

where

$$M_{mn}^{(\mathbf{k},\mathbf{k}+\mathbf{b})} = \langle u_{n\mathbf{k}} | u_{m\mathbf{k}+\mathbf{b}} \rangle, \tag{4.7}$$

are overlap matrices discussed in Chapter 2 and **b** are "mesh vectors" connecting each **k**-point to its nearest neighbors. The latter are chosen, together with a set of weights ω_b, in such a way as to satisfy the condition

$$\sum_{\mathbf{b}} \omega_b b_i b_j = \delta_{ij}. \tag{4.8}$$

A corresponding expression for $\tilde{\Omega}$, and a description of steepest-descent methods capable of minimizing Ω, were also given in Ref. [32].

In 1D the problem of finding maximally localized WFs has a simple solution. Consider a projected position operator $\hat{P}\hat{x}\hat{P}$, where $\hat{P}$ is the projector onto the occupied space. According to Nenciu [102] such operators are well-defined and they commute with lattice translations. If WFs are taken to be the eigenstates of the projected position operator $\hat{P}\hat{x}\hat{P}|\mathbf{0}n\rangle = \bar{x}_n|\mathbf{0}n\rangle$, then [103]

$$\langle \mathbf{R}m|\hat{x}|\mathbf{0}n\rangle = \langle \mathbf{R}m|\hat{P}\hat{x}\hat{P}|\mathbf{0}n\rangle = \bar{x}_n \delta(\mathbf{R})\delta_{mn} \tag{4.9}$$

and $\tilde{\Omega}$ vanishes identically. Therefore, they are just the desired maximally localized WFs. In dimensions higher than one the situation is much more complicated, since $\hat{P}\hat{x}\hat{P}$, $\hat{P}\hat{y}\hat{P}$ and $\hat{P}\hat{z}\hat{P}$ do not commute with each other, and it is impossible to make $\tilde{\Omega}$ vanish by finding a common set of eigenvectors. Instead one should search for the gauge that provides the best compromise between them and minimizes the gauge-dependent part of the spread functional. In this respect it is important to note that the maximally localized WFs in 2D or 3D are in general more delocalized in each of the directions then those that minimize the spread in only one given direction.

4.2 Hybrid Wannier functions

The simplicity of finding maximally localized WFs in 1D makes it tempting to consider hybrid (hermaphrodite) WFs [32, 104]. These are functions that are Wannier-like (localized) in one direction, but Bloch-like (delocalized) in all others,

$$|R_x k_y k_z n\rangle = \frac{L}{2\pi} \int_{-\pi/L}^{\pi/L} dk_x e^{-ik_x R_x} |\psi_{n\mathbf{k}}\rangle, \tag{4.10}$$

where L is the length of the lattice parameter in the x-direction, and $R_x = mL$, m being an integer. These functions provide a generalization of 1D WFs to higher dimensions.

In 1D there is a nice relation between the geometric quantities of Chapter 3 and WCCs. As shown by Blount [105]

$$\langle Rn|\hat{x}|0m\rangle = i\frac{L}{2\pi} \int_{BZ} e^{ikR} \langle u_{nk}|\partial_k|u_{mk}\rangle. \tag{4.11}$$

From this equation, using the fact that BZ in 1D is a closed loop, the 1D WCCs are nicely expressed in terms of the Berry phase [7]

$$\bar{x}_n = \langle 0n|\hat{x}|0n\rangle = i\frac{L}{2\pi} \oint A_{nn,x}(k) dk. \tag{4.12}$$

A gauge transformation of the form $|\tilde{u}_{nk}\rangle = e^{ikmL}|u_{nk}\rangle$ will result in a shift of the WCC by a lattice vector, i.e., $\tilde{\bar{x}}_n = \bar{x}_n + mL$. This is a consequence of the above-mentioned fact that the action of the lattice translation operator on the WF simply shifts it to a different unit cell. In this sense, the WCC of an isolated band is defined only modulo a lattice vector. Meanwhile, when dealing with an isolated group of bands, only the sum of WCCs is gauge-independent (up to a lattice vector), being related to the electric polarization [33] via

$$P = \frac{e}{L} \sum_{n=1}^{\mathcal{N}} \bar{x}_n, \tag{4.13}$$

while the individual $\bar{x}_n$ are gauge-dependent.

All of the above can be applied to hybrid WFs. The difference in this case is in the dependence of a hybrid WCC on k_y and k_z,

$$\bar{x}_n(k_y, k_z) = \langle 0 k_y k_z n | \hat{x} | 0 k_y k_z n \rangle. \tag{4.14}$$

It turns out that such a map of the hybrid WCCs can reveal certain information about the topology of the underlying system.

4.3 Chern numbers via hybrid WFs

Consider for simplicity a 2D insulator with a single occupied band. We can construct a hybrid WF for this band and obtain $\bar{x}(k_y)$ (or alternatively $\bar{y}(k_x)$). The position of the charge center at each given k_y is gauge-invariant modulo a lattice vector. Thus, if the gauge is continuous in k_y, we get a smooth continuous line evolving from $k_y = -\pi/L_y$ to $k_y = \pi/L_y$ as shown in Fig. 4.2.

This figure has a nice physical analogy. Consider that instead of a 2D system, we have a 1D system that depends on an external parameter k_y. Since the BZ is periodic, the evolution of k_y across the BZ realizes a periodic a cyclic process. We look at how a charge center moves in the 1D system during the cycle. The 1D system that we consider is periodic itself, so we can depict it as a circle. Thus, the motion of a hybrid WCC in the Fig. 4.2 illustrates the motion of a charge on the circumference during a cyclic process performed on the system (see the right side of the figure).

For a gauge that is smooth in k_y from $-\pi/L_y$ to π/L_y, the charge must be

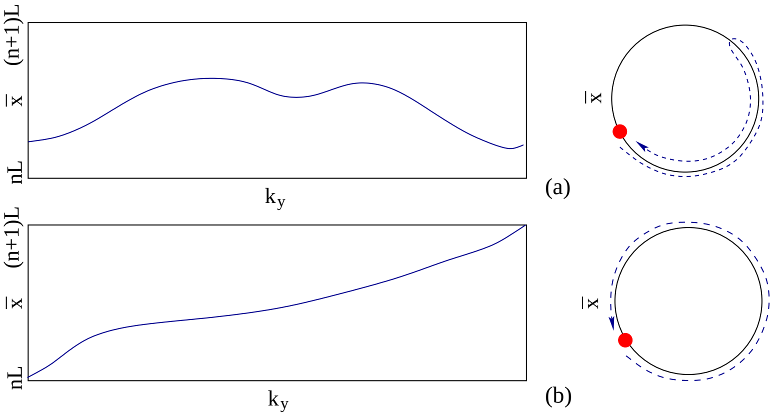

Figure 4.2: A sketch of $\bar{x}(k_y)$ dependence. Panel (a): $C = 0$. Panel (b): $C = 1$.

located at the same point on the circumference at the beginning and at the end of the cycle. However, it can arrive there by different routes. Different routes can be topologically distinct, depending on how many times (and in which direction) the charge winds around the circumference during the cycle. This winding number is equal to the charge (in units of e) pumped from one end of a long but finite 1D system to the other during the cycle.

The winding number turns out to be equal to the Chern number of the 2D insulator in question. Figure 4.2 (a) illustrates a typical hybrid WCC behavior [106] for the case of an ordinary insulator. The center comes back to its original location within the *same* unit cell, and the charge center of the hypothetical 1D system, associated with this insulator, does not wind. On the contrary, for an insulator with Chern number $C = 1$ the WCC shifts by one lattice vector, and the 1D charge center winds once around the system, as shown in Fig. 4.2 (b). The situation is quite general, and the shift of the hybrid WCC across the BZ is always equal to the Chern number, which is easy to prove as follows.

Since we assumed the gauge to be smooth in k_y, then for the shift of the hybrid WCC between $-\pi/L_y$ and π/L_y we get

$$\bar{x}(\pi/L_y) - \bar{x}(-\pi/L_y) = \frac{L_x}{2\pi}\left[\oint \mathcal{A}_x(k_y = \pi/L_y)dk_x - \oint \mathcal{A}_x(k_y = -\pi/L_y)dk_x\right], \tag{4.15}$$

and the mod-L_x ambiguity of the WCCs goes away.[1] The physical meaning of this is that we now consider the *change* in the position of the WCC, which does not depend on the unit cell the corresponding WF started from. The expression above can be viewed as taken over the boundary of the cylinder, e.g. the BZ is glued in the k_x direction, but not in k_y. Using the continuity of the gauge in k_y again, application of Gauss theorem leads to

$$\bar{x}(\pi/L_y) - \bar{x}(-\pi/L_y) = \frac{L_x}{2\pi}\int_{BZ} dk_x dk_y \left[\nabla_{\mathbf{k}} \times \boldsymbol{\mathcal{A}}\right]_z = \frac{L_x}{2\pi}\int dk_x dk_y \mathcal{F}_{xy}. \tag{4.16}$$

The integral above can be recognized from the definition of the Chern number (3.18). Thus, we conclude that indeed the shift of the hybrid WCC gives the Chern number of the system

$$\frac{1}{L_x}\left[\bar{x}(\pi/L_y) - \bar{x}(-\pi/L_y)\right] = C. \tag{4.17}$$

The generalization of this result to the multiband case is done in the same way and the result is

$$\frac{1}{L_x}\left[\sum_{n=1}^{\mathcal{N}} \bar{x}_n(\pi/L_y) - \bar{x}_n(-\pi/L_y)\right] = C. \tag{4.18}$$

It should be stressed that Eqs.(4.17-4.18) are valid only for a gauge that is smooth in k_y on the segment $[-\pi/L_y, \pi/L_y]$.

However, as was discussed in Sec. 2.1.2, k_y lives on a loop in the BZ. We may ask whether a gauge that is smooth on the whole loop rather than on an interval $[-\pi/L_y, \pi/L_y]$ exists. It turns out that when $C \neq 0$, such a gauge does not exist.

[1] Here we omit the band index, since the is only one band.

It has been proven that a construction of a complete set of well-localized 2D WF (not the hybrid ones) is impossible for a Chern insulator [36–38], meaning that a non-zero Chern number is an obstruction for choosing a smooth gauge globally in the BZ. In the above construction we implicitly took the gauge to be smooth on a BZ circle in the x-direction to construct hybrid WCCs,[2] so that the gauge has to be discontinuous in k_y. This gauge discontinuity is reflected in the fact that hybrid WCC do not come back to the original value after going across the BZ.

In the next chapter we will see that hybrid WCCs can also be used to compute the $\mathbb{Z}_2$ invariant. Unlike the case of Chern insulators, for $\mathbb{Z}_2$ insulators a globally smooth gauge does exist [36, 41], as discussed in Chapters 6 and 7. Given the importance of hybrid WFs in the following, we conclude this chapter with a recipe for constructing hybrid WFs on a discrete mesh of points starting from the random gauge. The gauge that we describe corresponds to maximal localization of a hybrid WF, and is sometimes referred to as a parallel transport gauge.

4.4 Parallel transport gauge

Let us discuss how to construct a parallel-transport gauge starting from a set of randomly chosen eigenstates of the Hamiltonian on a **k**-mesh [32]. In what follows we distinguish single-band and multiband parallel transport procedures. The general idea in both cases is to carry the Bloch states along a certain direction in the BZ (say, k_x) in such a way that they remain as “parallel” as possible to the previous states at all points. If the path is closed, the states might return to the initial point with some phase differences relative to the initial states, thus violating singlevaluedness. However, singlevaluedness of the wavefunction can be restored by spreading the extra phase uniformly along the path, as explained in more detail below. In this case, a closed loop is obtained when the state is

[2] Such a choice is always possible, since we can form the eigenstates of the position operator $\hat{x}$ as of the hybrid WFs, and obtain maximum localization in this direction as argued above.

transported by a reciprocal lattice vector $\mathbf{G}_x$. The generalization to an arbitrary direction should be obvious.

Consider a single isolated band $|u_{n\mathbf{k}}\rangle$. To carry the state to $k+\Delta k$ via parallel transport, the phase of the Bloch state at this new point should be chosen in such a way that the overlap $\langle u_{n\mathbf{k}}|u_{n,\mathbf{k}+\Delta k_x}\rangle$ is real and positive, so that the change in the state is orthogonal to the state itself. It is straightforward to implement this numerically. Consider a discrete uniform mesh of k-points $\{\mathbf{k}_j\}, j \in [1, N+1]$, where $\mathbf{k}_{j+1} = \mathbf{k}_j + \Delta k_x$ and $\mathbf{k}_{N+1} = \mathbf{k}_1 + \mathbf{G}_x$. The states $|\tilde{u}_{\mathbf{k}_j}\rangle$ at these points are obtained by a numerical diagonalization procedure and thus have random phases. At the initial point j=1 we set $|u'_{\mathbf{k}_1}\rangle = |\tilde{u}_{\mathbf{k}_1}\rangle$. Then at each subsequent $\mathbf{k}_{j+1}$ we let $\beta_{j+1} = \text{Im}\ \ln \langle \tilde{u}_{\mathbf{k}_{j+1}}|\, u'_{\mathbf{k}_j}\rangle$ and then apply the $\mathcal{U}(1)$ phase rotation

$$|u'_{\mathbf{k}_{j+1}}\rangle = e^{i\beta_{j+1}}|\tilde{u}_{\mathbf{k}_{j+1}}\rangle\,, \tag{4.19}$$

which makes $\langle u'_{\mathbf{k}_j}|u'_{\mathbf{k}_{j+1}}\rangle$ real and positive. Once this is done at each $\mathbf{k}$-point, the state at $\mathbf{k}_1$ differs from that at $\mathbf{k}_{N+1}$ by a phase factor $e^{i\phi}$, where ϕ is chosen on a particular branch, say $\phi \in (-\pi, \pi]$. ϕ is the Berry phase associated with the traversed path. Unless $\phi = 0$, periodicity in k_x is lost. To restore it, the extra phase should be spread uniformly along the string of $\mathbf{k}$-points, i.e.,

$$|u_{\mathbf{k}_j}\rangle = e^{-i\phi\mathbf{k}_j/2\pi}|u'_{\mathbf{k}_j}\rangle = e^{-i(j-1)\phi/N}|u'_{\mathbf{k}_j}\rangle, \tag{4.20}$$

where in the last equality the uniformity of the k-mesh was used.

In the multiband case one deals with the non-Abelian generalization of the Abelian Berry phase [63, 64]. We now consider an isolated set of $\mathcal{N}$ bands and describe parallel transport in the k_x-direction in the non-Abelian case [9, 32]. The parallel transport gauge is constructed by requiring that the overlap matrix

$$\tilde{M}_{mn}^{(\mathbf{k}_j,\mathbf{k}_{j+1})} = \langle \tilde{u}_{m\mathbf{k}_j}|\tilde{u}_{n\mathbf{k}_{j+1}}\rangle \tag{4.21}$$

must be Hermitian, with all positive eigenvalues, at each step. This is uniquely accomplished by means of the singular value decomposition in which an $\mathcal{N} \times \mathcal{N}$ matrix M is written in the form $M = V\Sigma W^\dagger$, where V and W are unitary and Σ is positive real diagonal. If the states at $\mathbf{k}_{j+1}$ are rotated by $\mathcal{U} = WV^\dagger$, i.e.,

$$|u'_{n\mathbf{k}_{j+1}}\rangle = \sum_m^{\mathcal{N}} \mathcal{U}_{mn}(\mathbf{k}_{j+1})|\tilde{u}_{m\mathbf{k}_{j+1}}\rangle, \tag{4.22}$$

the new overlap matrix $M'^{(\mathbf{k}_j,\mathbf{k}_{j+1})}_{mn}$ will be of the form $V\Sigma V^\dagger$, which is Hermitian with positive eigenvalues as desired. Repeating this procedure up to $j = N$, one obtains that the new states $|u'_{n\mathbf{k}_{N+1}}\rangle$ are related to the states $|u'_{n\mathbf{k}_1}\rangle$ by a unitary transformation Λ according to

$$|u'_{n\mathbf{k}_1}\rangle = e^{2\pi ix} \sum_m^{\mathcal{N}} \Lambda_{mn}|u'_{m\mathbf{k}_{N+1}}\rangle. \tag{4.23}$$

The eigenvalues of this matrix are of the form $\lambda_n = e^{-i\phi_n}$, where the phases $\phi_n = \operatorname{Im}\ln\lambda_n$ (again chosen according to some definite branch cut) are the analogs of the Abelian Berry phases.

To restore periodicity we follow the same trick as in the single-band case, but generalized to the matrix form. To do this one finds the unitary matrix R that diagonalizes Λ, and then rotates all states at all $\mathbf{k}_j$ by this same unitary R, so that the new states correspond to a diagonal Λ with its eigenvalues $\lambda_n = e^{i\phi_n}$ on the diagonal. Now it is straightforward to obtain periodicity by applying the graded phase twists

$$|u_{n\mathbf{k}_j}\rangle = e^{-i(j-1)\phi_j/N}|u'_{n\mathbf{k}_j}\rangle. \tag{4.24}$$

This results in a gauge that is smooth along k_x and $\mathbf{G}_x$-periodic.

Chapter 5

Computing topological invariants without inversion symmetry

We have seen above that the topology of the band structure can now be regarded as a fundamental characteristic of the electronic ground state for semiconductors and insulators. For this reason there is an obvious need in the methods that would allow for a routine computation of topological invariants of realistic materials in first-principles codes. In the materials with inversion symmetry such a method has been developed in Ref. ([31]), where the $\mathbb{Z}_2$ invariant of a centrosymmetric band structure was expressed as

$$(-1)^{\Delta} = \prod_{\mathbf{k}^*} \prod_{\alpha=1}^{\mathcal{N}/2} \zeta_{\alpha}, \tag{5.1}$$

where ζ_n is the parity of a Kramers pair,[1] and the products are taken over distinct occupied Kramers pairs and $\mathcal{T}$-invariant values of momentum in the BZ. This method is disarmingly simple, and it has been successfully implemented in *ab initio* calculations [70, 72]. However, a first-principles computation of the $\mathbb{Z}_2$ invariant for non-centrosymmetric materials remained problematic.

One possible approach is to apply the method of Fukui and Hatsugai [84], described in subsection 3.4.4. For the implementation of this approach, a gauge must be chosen on the boundary of half of the Brillouin zone (BZ) in such a way as to respect $\mathcal{T}$ symmetry, which involves acting with the $\mathcal{T}$ operator on one of

[1] Both states in the Kramers pair are guaranteed to have the same parity.

the states from each Kramers pair to construct the other. Although this method has been implemented in the *ab initio* framework [107–109], its implementation is basis-set dependent and involves the application of a unitary rotation to the computed eigenvectors when fixing the gauge, which may be tedious when there are many occupied bands and basis states.

Another existing method [70, 72] relies on the fact that the system will necessarily be in the $\mathbb{Z}_2$-even (normal) state in the absence of spin-orbit coupling. In this method, the strength of the spin-orbit coupling is artificially tuned from $\lambda_{SO} = 0$ (no spin-orbit coupling) to $\lambda_{SO} = 1$ (full spin-orbit coupling), and a closure of the band gap at some intermediate coupling strength is taken as evidence of an inverted band structure. However, a closure of the band gap in the course of tuning λ_{SO} to full strength is a necessary, but not a sufficient, condition for a topological phase transition. Therefore, in order to determine whether the system is really in the topologically nontrivial phase, a first-principles calculation of the surface states is carried out in order to count the number of Dirac cones at the surface of the candidate material. Such a calculation, although illustrative, is quite demanding in terms of computational resources.

In summary, existing methods have some shortcomings, and it would be very useful to develop a simple and effective method that would use the electronic wavefunctions, as obtained directly from the diagonalization procedure, to determine the desired topological indices.

In this chapter we develop a method for computing $\mathbb{Z}_2$ invariants that meets these criteria, and which is easy to implement in the context of *ab initio* code packages. The method is based on the concept of $\mathcal{T}$ polarization [27] (TRP), but implemented in such a way that a visual inspection of plotted curves is not required in order to obtain the topological indices. Instead, all the indices can be obtained directly as a result of an automated calculation. We describe the method, and then verify it using centrosymmetric Bi and Bi_2Se_3 as illustrative test examples before applying it to the more difficult cases of noncentrosymmetric GeTe and strained HgTe.

5.1 $\mathbb{Z}_2$ invariant via Wannier charge centers

In this section we review the notion of TRP and the definition of the $\mathbb{Z}_2$ invariant in terms of TRP derived in Ref. [27]. The definition arises by virtue of an analogy between a 2D $\mathcal{T}$-invariant insulator and a $\mathcal{T}$-symmetric pumping process in a 1D insulator. We further reformulate this definition in terms of Wannier charge centers, setting the stage for the numerical method discussed in the next section.

5.1.1 Review of time reversal polarization

Fu and Kane[27] considered a family of 1D bulk-gapped Hamiltonians $H(x)$ parametrized by a cyclic parameter t (i.e., $H[t+T] = H[t]$) subject to the constraint

$$H[-t] = \theta H[t]\theta^{-1}, \tag{5.2}$$

where θ is the $\mathcal{T}$ operator. This can be understood as an adiabatic pumping cycle, with t playing the role of time or pumping parameter. The constraint of Eq. (5.2) guarantees that the Hamiltonian $H(x)$ is $\mathcal{T}$-invariant at the points $t = 0$ and $t = T/2$, while the $\mathcal{T}$ symmetry is broken at intermediate parameter values. If we also limit ourselves to Hamiltonians having unit period, so that H is invariant under $x \to x + 1$, then the eigenstates may be represented by the cell periodic parts of the Bloch states, $|u_{nk}\rangle$. At $t = 0$ and $t = T/2$ the Hamiltonian is $\mathcal{T}$ invariant and the eigenstates come in Kramers pairs, being degenerate at $k = 0$ and $k = \pi$.

Since the system is periodic in both k and t, the $|u_{nk}\rangle$ functions are defined on a torus. Moreover, the system must also be physically invariant under a gauge transformation of the form (3.5) where $\mathcal{U}(\mathbf{k}) = \mathcal{U}(k,t)$ expresses the U($\mathcal{N}$) gauge freedom to choose $\mathcal{N}$ representatives of the occupied space at each (k,t) in the sense of Sec. 3.1. We adopt a gauge that is continuous on the half-torus $t \in [0, T/2]$ and that respects $\mathcal{T}$ symmetry at $t = 0$ and $T/2$ in the sense of Fu and Kane [27], i.e.,

$$\begin{aligned} |u^I_{\alpha,-k}\rangle &= -e^{i\chi_{\alpha k}}\theta|u^{II}_{\alpha k}\rangle, \\ |u^{II}_{\alpha,-k}\rangle &= e^{i\chi_{\alpha,-k}}\theta|u^{I}_{\alpha k}\rangle. \end{aligned} \tag{5.3}$$

Here the occupied states $n = 1, ..., \mathcal{N}$ have been relabeled in terms of pairs $\alpha = 1, ..., \mathcal{N}/2$ and elements I and II within each pair. Note that Eq. (5.3) is a property which is not preserved by an arbitrary $\mathcal{U}(\mathcal{N})$ transformation. It allows the Berry connection

$$\mathcal{A}(k) = i\sum_n \langle u_{nk}|\partial_k|u_{nk}\rangle \tag{5.4}$$

to be decomposed as

$$\mathcal{A}(k) = \mathcal{A}^I(k) + \mathcal{A}^{II}(k) \tag{5.5}$$

where

$$\mathcal{A}^S(k) = i\sum_\alpha \langle u^S_{\alpha k}|\partial_k|u^S_{\alpha k}\rangle \tag{5.6}$$

and $S = I, II$. Having chosen a gauge that obeys these conventions at $t = 0$ and $T/2$ and evolves smoothly for intermediate t, [2] the "partial polarizations" [27]

$$P^S_\rho = \frac{1}{2\pi}\oint dk\mathcal{A}^S(k) \tag{5.7}$$

[2] Since we do not constrain the gauge of the 1D system to obey any particular symmetries at intermediate t, it is always possible to perform the unitary mixing at intermediate t in such a way that the pair of "bands" belonging to the same α at t=0 also belong to the same α at t=$T/2$.

can be defined such that their sum is the total charge polarization [10]

$$P_\rho = \frac{1}{2\pi}\oint dk\mathcal{A}(k) = P_\rho^I + P_\rho^{II}. \tag{5.8}$$

Note that the total polarization is defined only modulo an integer (the quantum of polarization) under a general U($\mathcal{N}$) gauge transformation, while the "partial polarization" is not gauge invariant at all. A quantity that *is* gauge-invariant is the change in total polarization during the cyclic adiabatic evolution of the Hamiltonian, and using Eq. (5.2) it follows that

$$P_\rho(T) - P_\rho(0) = C \tag{5.9}$$

where C is the first Chern number, an integer topological invariant corresponding to the number of electrons pumped through the system in one cycle of the pumping process [12]. For a $\mathcal{T}$-invariant pump that satisfies the conditions of Eq. (5.2), C must be zero.

In order to describe the $\mathbb{Z}_2$ invariant of a $\mathcal{T}$-symmetric system in a similar fashion, the "time reversal polarization" was introduced as [27]

$$P_\theta = P_\rho^I - P_\rho^{II}. \tag{5.10}$$

Then the integer $\mathbb{Z}_2$ invariant can be written as

$$\Delta = P_\theta(T/2) - P_\theta(0) \mod 2. \tag{5.11}$$

To summarize, the $\mathbb{Z}_2$ invariant is well defined via Eq. (5.11) when the gauge respects $\mathcal{T}$-symmetry at $t = 0$ and $T/2$ and is continuous on the torus between these two parameter values. Note, however, that while such a gauge choice is possible on the half-torus even for the $\mathbb{Z}_2$-odd case (Δ=1), it can only be extended to cover the full torus continuously in the $\mathbb{Z}_2$-even case (Δ=0) [27, 40, 110].

5.1.2 Formulation in terms of Wannier charge centers

Let us now rewrite Eq. (5.11) in terms of the Wannier charge centers (WCCs). In the present consideration the definition (4.1) of the Wannier functions (WFs) is written as

$$|Rn\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} dk e^{-ik(R-x)}|u_{nk}\rangle. \tag{5.12}$$

The WCC $\bar{x}_n$ is defined as the expectation value $\bar{x}_n = \langle 0n|\hat{x}|0n\rangle$ of the position operator $\hat{x}$ in the state $|0n\rangle$ corresponding to one of the WFs in the home unit cell $R = 0$ or, equivalently [7, 10] (see section 4.2 for details)

$$\bar{x}_n = \frac{i}{2\pi}\int_{-\pi}^{\pi} dk \langle u_{nk}|\partial_k|u_{nk}\rangle. \tag{5.13}$$

As was discussed in the previous chapter, the sum of all WCCs is a gauge independent quantity defined modulo a lattice vector, i.e. $\mod 1$ in our notations [10]. Individual $\bar{x}_n$, however, apart from being defined only $\mod 1$, are gauge-dependent. For the present purposes we adopt the gauge of Eq. (5.3) and construct WFs $|R\alpha, S\rangle$ by inserting $|u^S_{\alpha k}\rangle$ into the definition of Eq. (5.12). In this gauge

$$\bar{x}^I_\alpha = \bar{x}^{II}_\alpha \mod 1, \tag{5.14}$$

as follows from Eqs. (5.3) and (5.13) and use of the continuity condition $\chi_{\alpha,-\pi} = \chi_{\alpha,\pi} + 2\pi m$, where m is an integer. Since we have also insisted on the gauge being continuous for $t \in [0, T/2]$, it is possible to follow the evolution of each WCC during the half-cycle. Taking into account that $\sum_\alpha \bar{x}^s_\alpha = (1/2\pi)\oint_{\mathrm{BZ}} \mathcal{A}^S$ for $S = I, II$, Eq. (5.11) yields

$$\Delta = \sum_\alpha \left[\bar{x}^I_\alpha(T/2) - \bar{x}^{II}_\alpha(T/2)\right] - \sum_\alpha \left[\bar{x}^I_\alpha(0) - \bar{x}^{II}_\alpha(0)\right]. \tag{5.15}$$

Since the gauge is assumed to be smooth, the evolution of the charge centers must also be smooth. Being defined in this way, Δ is clearly a mod-2 quantity, and as shown in Ref. [27] it represents the desired $\mathbb{Z}_2$ invariant.

However, if the gauge breaks $\mathcal{T}$ symmetry or it is not continuous in the half-cycle, Eq. (5.15) no longer defines a topological invariant. A discontinuity in the gauge in the process of the half cycle can change Δ by 1, so the mod-2 property is lost. Breaking $\mathcal{T}$ in the gauge choice means that the corresponding centers are not necessarily degenerate at $t = 0$ and $t = T/2$. In fact, Δ can even take non-integer values in this case [40].

The above argument implies that in order to compute the $\mathbb{Z}_2$ invariant via Eq. (5.15), one needs a gauge that satisfies both $\mathcal{T}$-invariance and continuity on the half-torus. We now argue that the gauge that corresponds to 1D maximally localized WFs at each t has the desired properties, as long as these WFs are chosen to evolve smoothly as a function of t. According to Sec. 4.1, the maximally localized WFs in 1D are eigenstates of the position operator $\hat{x}$ in the band subspace [32, 103]. Since this operator commutes with θ, its eigenvalues will be doubly degenerate and its eigenstates will come in Kramers pairs at $t = 0$ and $T/2$.

The continuity of the gauge in k is obtained by carrying out a multiband parallel transport of Sec. 4.4 along the BZ [32]. Eq. (4.23) relates the states $|\psi_{nk}\rangle$ at $k = 2\pi$, obtained via parallel transport, to those at $k = 0$ by a unitary rotation Λ, whose eigenvalues $\lambda_n = e^{-i\bar{x}_n}$ give the 1D maximally-localized WCCs $\bar{x}_n$. The corresponding eigenvectors can be used to define a gauge that is continuous in k for a given value of t. The continuity vs. t on the half-torus is achieved by tracing the evolution of the WCCs $\bar{x}_n$ as a function of t, with the n'th state of the gauge constructed from the eigenvectors associated with the n'th smoothly evolving WCC $\bar{x}_n(t)$.

Having established a particular gauge choice in which Eqs. (5.11) and (5.15) are valid, it is straightforward in principle to obtain the $\mathbb{Z}_2$ invariant. Indeed, Eq. (5.15) implies that the $\mathbb{Z}_2$ invariant can be determined simply by testing whether the WCCs change partners when tracked continuously from t=0 to $T/2$. This is the essence of our approach. We stress that no explicit construction of a smooth gauge on the half-torus is necessary; we simply track the evolution of the WCCs on the half-torus.

In practice, when working on a discrete mesh of t values when many bands are present, it may not be entirely straightforward to enforce the continuity with respect to t. In the next section we present a simple and automatic numerical procedure that is robust in this respect, and use it to illustrate the calculation of the $\mathbb{Z}_2$ invariants for several materials of interest.

5.2 Numerical Implementation

The method outlined above, in which the WCCs obtained with the 1D maximally-localized gauge are used to compute the $\mathbb{Z}_2$ invariant via Eq. (5.15), can be implemented by plotting the WCCs at each point on the t mesh and then visually tracking the evolution of each WCC, as we describe next in Sec. 5.3.1. However, we find that a more straightforward and more easily automated approach is to track the *largest gap* in the spectrum of WCCs instead. This gives rise to our proposed method, which is described in Sec. 5.3.2.

5.3 Numerical Implementation

The method outlined above, in which the WCCs obtained with the 1D maximally-localized gauge are used to compute the $\mathbb{Z}_2$ invariant via Eq. (5.15), can be implemented by plotting the WCCs at each point on the t mesh and then visually tracking the evolution of each WCC, as we describe next in Sec. 5.3.1. However, we find that a more straightforward and more easily automated approach is to

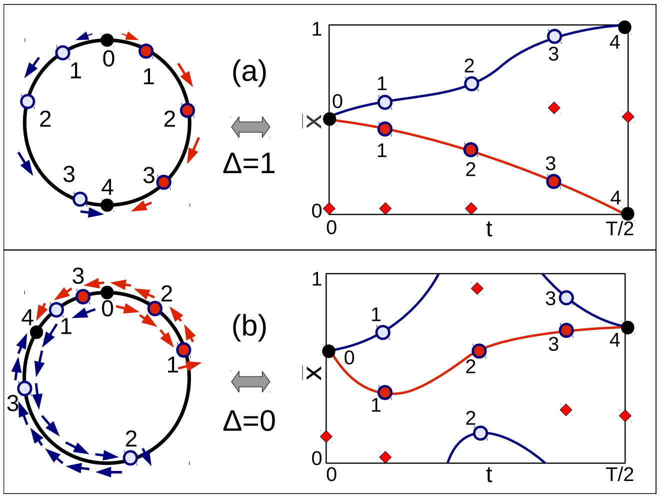

Figure 5.1: Sketch of evolution of Wannier charge centers (WCCs) $\bar{x}$ vs. time t during an adiabatic pumping process. Regarding $\bar{x} \in [0,1]$ as a unit circle and $t \in [0,T/2]$ as a line segment, the cylindrical $(\bar{x},t)$ manifold is represented via a sequence of circular cross sections at left, or as an unwrapped cylinder at right. Each red rhombus marks the middle of the largest gap between WCCs at given t. (a) $\mathbb{Z}_2$ insulator; WCCs wind around the cylinder. (b) Normal insulator; WCCs reconnect without wrapping the cylinder.

track the *largest gap* in the spectrum of WCCs instead. This gives rise to our proposed method, which is described in Sec. 5.3.2.

5.3.1 Tracking WCC locations

Let us first interpret Eq. (5.15) in terms of the winding of the WCCs around the BZ during the half-cycle $t \in [0,T/2]$. Since the WCCs are defined modulo 1, one can imagine the $\bar{x}_n$ living on a circle of unit circumference, as illustrated in the left panels of Fig. 5.1. During the pumping process, the WCCs migrate along this circle. The system will be in the $\mathbb{Z}_2$-odd state (δ=1) if and only if the WCCs reconnect after the half cycle in such a way as to wrap the unit circle an odd number of times.

Consider, for example, the case of only two occupied bands, as sketched in Fig. 5.1. The top panel shows the $\mathbb{Z}_2$-odd case; the blue and green arrows show the evolution of the first and second WCC from t_0 $(=0)$ to t_4 $(=T/2)$, and they meet in such a way that the unit circle is wrapped exactly once. Correspondingly, as shown in the right-hand part of the figure, the WCCs "exchange partners" during the pumping process (i.e., two bands belonging to the same Kramers pair at $t = 0$ do not rejoin at $t = T/2$) [27]. For the $\mathbb{Z}_2$-even case shown in the bottom panel, by contrast, the unit circle is wrapped zero times, and no such exchange of partners occurs.

If one has access to the continuous evolution of the WCCs vs. t, as shown by the solid blue and green curves in Fig. 5.1, this method works in principle for an arbitrary number of occupied bands (i.e., WFs per unit cell). An illustrative example with many bands appears in Fig. (1) of Ref. [111]. Either the "bands" $\bar{x}_n$ exchange partners in going from $t = 0$ to $t = T/2$ ($\phi = 0$ to $\phi = \pi$ in their notation), or they do not, implying $\mathbb{Z}_2$ odd or even respectively. Equivalently, one can draw an arbitrary continuous curve starting within a gap at $t = 0$ and ending within a gap at $t = T/2$; the system is $\mathbb{Z}_2$-odd if this curve crosses the WCC bands an odd number of times, or $\mathbb{Z}_2$-even otherwise.

In practice, however, one will typically have the WCC values only on a discrete mesh of t points, in which case the connectivity can be far from obvious. Certainly one cannot simply make the arbitrary branch cut choice $\bar{x}_n \in [0, 1]$, sort the $\bar{x}_n$ in increasing order, and use the resulting indices to define the paths of the WCCs. This would, for example, give an incorrect evolution from t_1 to t_2 in Fig. 5.1(b), since one WCC passes through the branch cut in this interval, apparently jumping discontinuously from the "top" to the "bottom" of the unwrapped cylinder at right. (A similar jump happens again near t_3.)

One possible approach is that of Ref. [111] mentioned above, i.e., to increase the t mesh density until, by visual inspection, the connectivity becomes obvious. However, this becomes prohibitively expensive in the first-principles context, since a calculation of many (typically 10-30) bands would have to be done on an extremely fine mesh of t points. It is typical for some of the WCCs to cluster rather closely together during part of the evolution in t; if this clustering happens near the artificial branch cut, it can become very difficult to determine the connectivity from one t to the next, even if a rather dense mesh of t values is used. Moreover, an algorithm of this kind is difficult to automate. For these reasons, we find that the direct approach of plotting the evolution of the WCCs is not a very satisfactory algorithm for obtaining the topological indices, at least in the case of a large number of occupied bands.

5.3.2 Tracking gaps in the WCC spectrum

Here we propose a simple procedure that overcomes the above obstacles, allowing the $\mathbb{Z}_2$ invariant to be computed in a straightforward fashion. The main idea is to concentrate on the *largest gap between WCCs*, instead of on the individual WCCs themselves. As explained above and illustrated by the red dashed curve in Fig. 5.1, the path following the largest gap in $\bar{x}_n$ values (with vertical excursions at critical values of t) crosses the $\bar{x}_n$ bands a number of times that is equal, mod 2, to the $\mathbb{Z}_2$ invariant. Our approach, in which we choose this path as an especially suitable one for discretizing, can be implemented without reference to any branch cut in the determination of the $\bar{x}_n$, allowing the $\mathbb{Z}_2$ invariant to be determined from the flow of WCCs on the cylindrical $(\bar{x}, t)$ manifold directly.

As in Fig. 5.1, we again consider a set of M circular sections of the cylinder that correspond to the pumping parameter values $t^{(m)} = T(m-1)/2M$, where $m \in [0, M]$. At each t_m we define $z^{(m)}$ to be the center of the largest gap between two adjacent WCCs on the circle. (If two gaps are of equal size, either can be chosen arbitrarily.) For definiteness we choose $z^{(m)} \in [0, 1)$, but as we shall see shortly, the branch choice is immaterial. In the continuous limit $M \to \infty$, $z(t)$ takes the form of a series of path segments on the surface of the cylinder, with discontinuous jumps in the $\bar{x}$ direction at certain critical parameter values t_j. Our algorithm consists in counting the number of WCCs jumped over at each t_j, and summing them all mod 2. As becomes clear from an inspection of Fig. 5.1 and similar examples of increasing complexity, the WCCs exchange partners during the evolution from t=0 to $T/2$ only if this sum is odd, so that this sum determines the $\mathbb{Z}_2$ invariant of the system.

The approach generalizes easily to the case of discrete $z^{(m)}$. Let Δ_m be the number of WCCs $\bar{x}_n^{(m+1)}$ that appear between gap centers $z^{(m)}$ and $z^{(m+1)}$, mod 2. As we shall see below, this can be computed in a manner that is independent of the branch cut choices used to determine the $\bar{x}_n^m$ and $z^{(m)}$. Then the overall $\mathbb{Z}_2$ invariant is just

$$\Delta = \sum_{m=0}^{M} \Delta_m \quad \text{mod } 2. \tag{5.16}$$

This argument is illustrated in the right-hand panels of Fig. 5.1 for the two band-case and $M = 4$. The rectangles represent the surface of the cylinder in the parameter space, and should be regarded as glued along the longer sides. The circles correspond to $\bar{x}_n^{(m)}$ values, while each red rhombus represents the center $z^{(m)}$ of the largest gap between $\bar{x}_n^{(m)}$ values. In Fig. 5.1(a) there is one jump that occurs between m=2 and m=3, in which one WCC is jumped over; thus, $\Delta_m = 0$ except for $\Delta_2 = 1$, giving Δ=1. In Fig. 5.1(b), on the other hand, there are two jumps, once between m=1 and m=2 and again between m=2 and m=3, so that $\Delta_1 = \Delta_2 = 1$ and $\Delta = 0$ (mod 2).

We now show how the Δ_m can be computed straightforwardly in a manner that is insensitive to the branch-cut choices made in determining the $\bar{x}_n^m$ and $z^{(m)}$. We use the fact that the directed area of a triangle defined by angles ϕ_1, ϕ_2, and

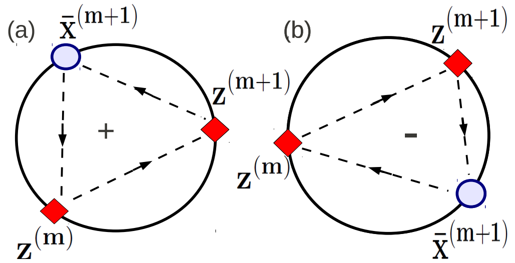

Figure 5.2: Sketch illustrating the method used to determine whether $\bar{x}_n^{(m+1)}$ lies between $z^{(m)}$ and $z^{(m+1)}$ in the counterclockwise sense when mapped onto the complex unit circle. (a) Yes, since the directed area of the triangle is positive. (b) No, since it is negative.

ϕ_3 on the unit circle is [3]

$$g(\phi_1, \phi_2, \phi_3) = \sin(\phi_2 - \phi_1) + \sin(\phi_3 - \phi_2) + \sin(\phi_1 - \phi_3). \tag{5.17}$$

Therefore the sign of $g(\phi_1, \phi_2, \phi_3)$ tells us whether or not ϕ_3 lies "between" ϕ_1 and ϕ_2 in the sense of counterclockwise rotation. Identifying $\phi_1 = 2\pi z^{(m)}$, $\phi_2 = 2\pi z^{(m+1)}$ and $\phi_3 = 2\pi \bar{x}_n^{(m+1)}$, as in Fig. 5.2, it follows that

$$(-1)^{\Delta_m} = \prod_{n=1}^{\mathcal{N}} \text{sgn}\left[g(2\pi z^{(m)}, 2\pi z^{(m+1)}, 2\pi \bar{x}_n^{(m+1)})\right], \tag{5.18}$$

where sgn(x) is the sign function. The Δ_m defined in this way is precisely the needed count of WCCs jumped over, mod 2, in evolving from m to $m+1$.

As a last detail, we discuss the case of possible degeneracies between the three arguments of $g(\phi_1, \phi_2, \phi_3)$. First, note that $z^{(m+1)} = \bar{x}_n^{(m+1)}$ is impossible, since $z^{(m+1)}$ is by definition in a gap between $\bar{x}_n^{(m+1)}$ values. If the mesh spacing in t is fine enough, then by continuity we expect that $z^{(m)} = \bar{x}_n^{(m+1)}$ will also be unlikely. It is recommended to test whether these values ever approach within a

[3] This follows from the fact that the directed area of the triangle defined by vertices z_j in the complex plane is $\text{Im}[z_1^* z_2 + z_2^* z_3 + z_3^* z_1]$; specializing this to $z_j = \exp(i\phi_j)$ yields Eq. (5.17).

threshold distance, and restart the algorithm with a finer t mesh if such a case is encountered; two cases of this kind are discussed later in Sec. 5.4. Finally, it can happen that $z^{(m)} = z^{(m+1)}$. In this case, the signum function (which technically assigns value 0 to argument 0) should be replaced in Eq. (5.18) by a function that returns s whenever $z^{(m)} = z^{(m+1)}$, where s is chosen once and for all to be either $+1$ or -1. Since the same degeneracy appears in every term of the product over $\mathcal{N}$ factors in Eq. (5.18), where $\mathcal{N}$ is even, the choice of s is arbitrary as long as it is applied consistently.

The above-described algorithm, based on Eqs. (5.16-5.18), constitutes one of the principal results of the present work. The implementation of this algorithm is straightforward, and allows for an efficient and robust determination of the $\mathbb{Z}_2$ invariant even when many bands are present, and even for only moderately fine mesh spacings. In Sec. 5.4, we will demonstrate the successful application of this approach to the calculation of the strong and weak topological indices of some real materials.

5.3.3 Application to 2D and 3D $\mathcal{T}$-invariant insulators

As pointed out in Ref. [27], the pumping process discussed above for a 1D system is the direct analogue of a 2D $\mathcal{T}$-invariant insulator, i.e., one whose Hamiltonian is subject to the condition $H(-\mathbf{k}) = \theta^{-1}H(\mathbf{k})\theta$. To see this, let $\mathbf{k} = \sum_i k_i\mathbf{b}_i/2\pi$, where $\mathbf{b}_1$ and $\mathbf{b}_2$ have been chosen as primitive reciprocal lattice vectors. Then we can let k_1 and k_2 play the roles of k and t respectively. Just as $H(k,t)$ displays $\mathcal{T}$ symmetry of $H(x)$ at t=0 and $T/2$, so $H(k_1, k_2)$, regarded as the Hamiltonian $H(x_1)$ of a fictitious 1D system for given k_2, is $\mathcal{T}$-invariant at $k_2 = 0$ and π. The Wannier functions of the effective 1D system can be understood as hybrid Wannier functions of Sec. 4.2 that have been Fourier transformed from k space to r space only in direction 1, while remaining extended in direction 2. The topological $\mathbb{Z}_2$ invariant of the 2D system can therefore be determined straightforwardly by applying the approach outlined above.

A complete topological classification in 3D, outlined in Sec. 3.4.5, is given by the index $\nu_0; (\nu_1\nu_2\nu_3)$. These indices can be obtained by applying our analysis to each of these six faces in the 3D Brillouin zone. Note that in general, this determines the strong index ν_0 with some redundancy, providing a check on the internal consistency of the method. However, symmetry considerations often play a role. For systems having a 3-fold symmetry axis, for example, one typically needs to compute the $\mathbb{Z}_2$ index on only two faces, as we shall see below.

5.4 Application to real materials

In this section we discuss the application of the above-described method to real materials. First, we illustrate the validity of the approach for centrosymmetric Bi and Bi_2Se_3, where weak and strong indices may alternatively be computed directly from the parities of the occupied Kramers pairs at the eight $\mathcal{T}$-invariant momenta [31]. We then apply the method to noncentrosymmetric crystals of GeTe and strained HgTe, showing that the first is a trivial insulator, while the latter is a strong topological insulator (TI) under both positive and negative strains along [001] and under positive strain along [111].

The calculations were carried out in the framework of density-functional theory [44, 45] using the local-density approximation with the exchange and correlation parametrized as in Ref. [50]. We used HGH pseudopotentials [58] with semicore $5d$-states included for Hg, while for all other elements only the s and p valence electrons were explicitly included. The calculations were carried out using the Abinit code package [56, 57] with a $10 \times 10 \times 10$ $\mathbf{k}$-mesh for the self-consistent field calculations and a 140 Ry planewave cutoff. The spin-orbit interaction was included in the calculation via the HGH pseudopotentials. Note that the overlap

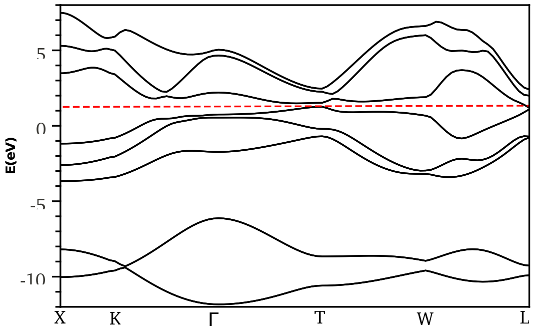

Figure 5.3: Band structure of Bi along high symmetry lines in the BZ. Fermi level is shown in red to illustrate the semimetallic nature of Bi.

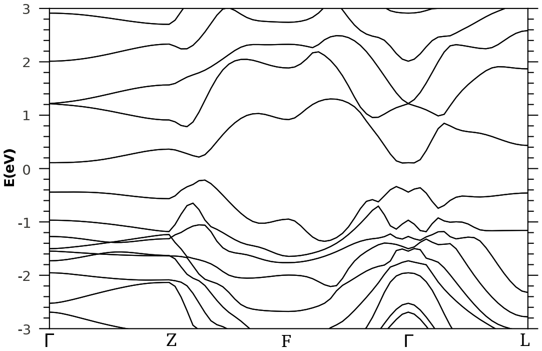

Figure 5.4: Band structure of Bi_2Se_3 along high symmetry lines in the BZ.

matrices $M_{mn}^{(k_j,k_{j+1})}$ defined in Sec. 5.1.2, are the same as those needed for the calculation of the electric polarization [10] or the construction of maximally-localized Wannier functions [32], and are thus readily available in many standard *ab initio* code packages including ABINIT.

5.4.1 Centrosymmetric materials

We start by illustrating the method with the examples of Bi and Bi_2Se_3. The band structures of these materials are shown in Fig. 5.3 and 5.4 respectively. Although Bi is a semimetal, its ten lowest-lying valence bands are separated from higher ones by an energy gap everywhere in the BZ, so in this case the topological indices describe the topological character of a particular group of bands. Since this is not the occupied subspace of an insulator, these topological indices are not "physical," but it is still of interest to compute them and compare with methods based on the parity eigenvalues [31]. According to the latter approach, the group of ten lowest-lying bands of Bi was shown to be topologically trivial [31]. Bi_2Se_3, on the other hand, is a true insulator, and the parity approach demonstrated that it is a strong topological insulator [70].

Bi and Bi_2Se_3 both belong to the rhombohedral space group $R\bar{3}m$ (#166), which has a 3-fold rotational axis. Thus, it is enough to compute only one weak $\mathbb{Z}_2$ index, say for $n_1 = 1$, since all three of them are equal by symmetry. To get the strong index, one just needs to compute just one more of the $\mathbb{Z}_2$ invariants, say for $n_1{=}0$.

Our results for Bi, obtained with the lattice parameters used in previous studies [112], are presented in Fig. 5.5. Panels (a) and (b) show the determination of the $\mathbb{Z}_2$ invariant at $n_1{=}0$ and $n_1{=}1$ respectively, with k_2 treated as the pumping parameter (like t) for an effective 1D system with wavevector k_3. The k_2 axis was initially discretized into ten equal intervals ($m = 1, ..., 10$) running from 0 to π, but for reasons discussed below an extra point (number 10 on the horizontal axis of the plot) was inserted midway in the last segment to make a total of eleven m values in Panel (b). As noted above, we are treating a group of ten valence bands labeled by n, so we have an array of WCC values $\bar{x}_n^{(m)}$ whose values are indicated by the black circles in the plot. These form Kramers pairs at $k_2{=}0$ and π, but not elsewhere. Each red rhombus indicates the center $z^{(m)}$ of the largest gap between adjacent $\bar{x}_n^{(m)}$ values, as discussed in Sec. 5.3.

Looking first at Fig. 5.5(a), we see that the gap center jumps over one WCC at $m{=}1$, and then over three WCCs at $m{=}7$, for a total of four, which is even. In Fig. 5.5(b) we get a total of 2+7+3+4 = 16 jumps, which is again even. The visual determinations of the number of jumped bands is confirmed by the application of

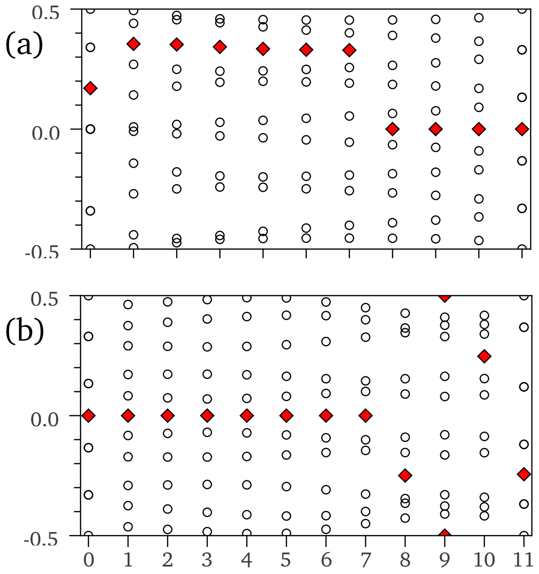

Figure 5.5: Evolution of Bi WCCs $\bar{x}_n$ (circles) in the r_3 direction vs. k_2 at (a) k_1=0; (b) k_1=π. Red rhombus marks midpoint of largest gap. k_2 is sampled in ten equal increments from 0 to π, except that an extra point is inserted midway in the last segment in panel (b) (see text).

the automated procedure of Eqs. (5.16-5.18). Thus, both $\mathbb{Z}_2$ indices are 0, and the 3D index is $0; (000)$, indicating a normal band topology as anticipated [31, 84].

We now discuss the above-mentioned insertion of one extra k_2 point in Fig. 5.5(b). This was necessary because the gap center $z^{(9)}$ at $k_2 = 0.9\pi$ had almost the same value as one of the WCC values at $k_2 = \pi$ (now labeled as '11' on the horizontal axis), making it ambiguous whether or not that x_n value should be counted as one of the ones that has been jumped over. To resolve this difficulty, we included an extra step at $k_2 = 0.95\pi$ (now labeled as '10' on the horizontal axis). The reason for the fast motion of the WCC in this case is that the minimum gap to the next higher (eleventh) band becomes rather small near $k_2 = \pi$.

Note that the detection of this kind of problem does not have to be done by visual inspection, but can be automated in the context of Eqs. (5.16-5.18). As already mentioned in Sec. 5.3.2, we simply test whether any $\bar{x}_n^{(m+1)}$ approaches within a certain threshold of $z^{(m)}$ (mod 1); if so, we flag the interval in question

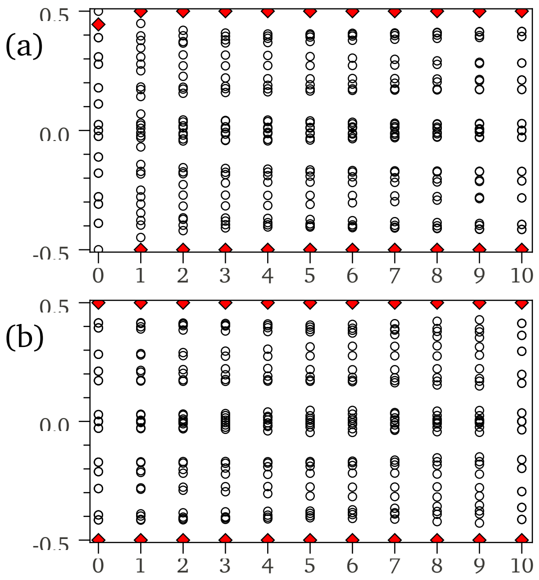

Figure 5.6: Evolution of Bi_2Se_3 WCCs $\bar{x}_n$ (circles) in the r_3 direction vs. k_2 at (a) k_1=0; (b) k_1=π. Red rhombus marks midpoint of largest gap. k_2 is sampled in ten equal increments from 0 to π.

for replacement by a finer mesh. Still, it is recommended to choose a mesh that is fine enough so that this threshold is rarely encountered, with a finer mesh recommended in cases where the minimum band gap is small.[4]

The analysis of the same n_1=0 and n_1=1 faces for the 28 WCCs of Bi_2Se_3 is illustrated in Fig. 5.6. The experimental lattice parameters [113] were used. Here there are no jumps over WCCs except for a single one in the very first step in the top panel (n_1=0). It follows that the topological index is 1; (000), in accord with previous studies [70].

5.4.2 Noncentrosymmetric materials

We now proceed to systems without inversion symmetry, which are the principal targets of our method since an analysis based on parity eigenvalues is not possible.

[4] In the vicinity of a small gap, it is also advisable to reduce the mesh spacing along the k-point strings used for the parallel transport construction.

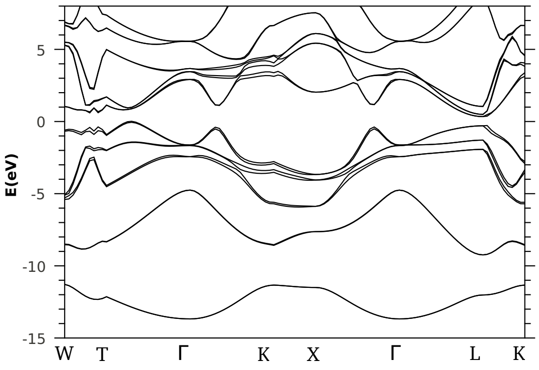

Figure 5.7: Band structure of GeTe along high symmetry lines in the BZ.

GeTe belongs to the rhombohedral $R3m$ space group (#160) and has no inversion symmetry (see Fig. 5.7 for the band structure), although like Bi and Bi_2Se_3 it has a 3-fold rotational symmetry, so that only two reciprocal-space faces have to be studied. The experimental lattice parameters [114] were used, and the evolution of the 10 WCCs is presented in Fig. 5.8 following similar conventions as for Bi and Bi_2Se_3. For both faces Eq. (5.18) gives a trivial $\mathbb{Z}_2$ index, with the center of the largest gap making no jumps, so that GeTe is in the topologically trivial state $0;(000)$. This result could have been anticipated from the fact that the spin-orbit interaction in GeTe is weak, as reflected in the approximate pairwise degeneracy of the WCCs throughout the evolution.

Finally, let us consider the more interesting case of epitaxially strained HgTe. In the absence of strain this is a zero-band-gap material. Any anisotropic strain breaks the four-fold symmetry at Γ, making it possible that the gap might open. Based on an adiabatic continuity argument, HgTe was predicted to be a strong TI under compressive strain in the [001] direction [31]. This was later verified with tight-binding calculations [25, 115]. Application of our approach to HgTe under uniaxial strain also confirms that HgTe is a strong topological insulator, with index $1;(000)$, under both positive and negative [31] 2% epitaxial strains along the [001] direction (not shown). This means that although the positive-strain and

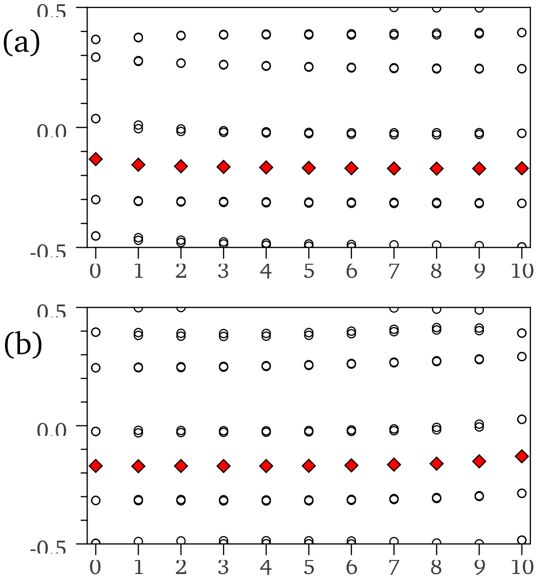

Figure 5.8: Evolution of GeTe WCCs $\bar{x}_n$ (circles) in the r_3 direction vs. k_2 at (a) k_1=0; (b) k_1=π. Red rhombus marks midpoint of largest gap. k_2 is sampled in ten equal increments from 0 to π.

negative-strain states are separated by a gap closure at zero strain, there is no topological phase transition associated with this gap closure.

We also studied epitaxial strains in the [111] direction. Under compressive strains of -2% and -5% the system becomes metallic and the direct band gap vanishes, so that no topological index can be associated with the occupied space. Under tensile strain of $+2\%$ we find that HgTe becomes a narrow-gap semiconductor with an indirect energy gap of $E_g = 0.054\,\text{eV}$, while for $+5\%$ strain it becomes metallic. Even at $+5\%$ strain, however, the lowest 18 bands remain separated from higher ones by an energy gap at all $\mathbf{k}$, so that, as for Bi, one can still assign a topological index to this isolated group of bands. The computed band structures for both cases are illustrated in Fig. 5.9 along lines connecting the high symmetry points of the undistorted FCC structure.

The space group of [111]-strained HgTe is rhombohedral $R3m$ (#166), the same as for GeTe, so that again only two $\mathbb{Z}_2$ indices need to be calculated. The

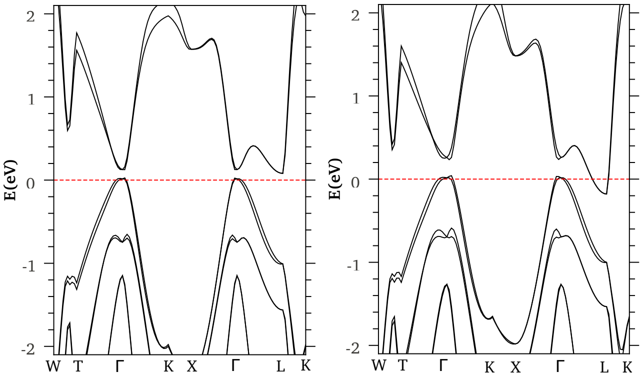

Figure 5.9: Band structure along high-symmetry lines of the undistorted FCC structure for HgTe under tensile epitaxial strain in the [111] direction. (a) +2% epitaxial strain. (b) +5% epitaxial strain.

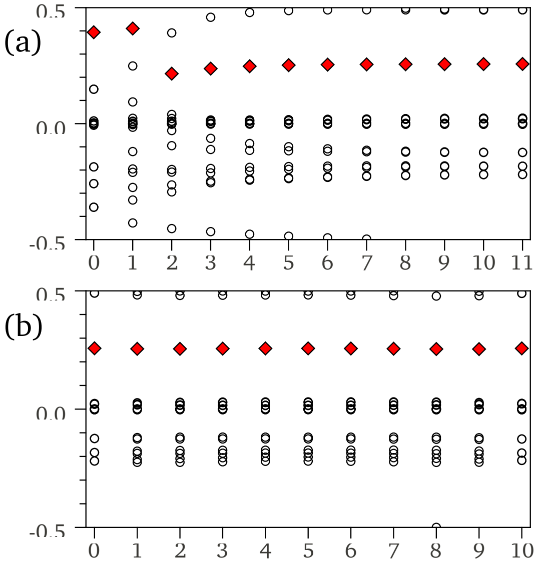

Figure 5.10: Evolution of WCCs for HgTe under +2% epitaxial strain in the [111] direction. WCCs $\bar{x}_n$ (circles) in the r_3 direction are plotted vs. k_2 at (a) k_1=0; (b) k_1=π. Red rhombus marks midpoint of largest gap. k_2 is sampled in ten equal increments from 0 to π, except that an extra point is inserted midway in the first segment in panel (a) (see text).

results of our WCC analysis for the case of +2% strain are shown in Fig. 5.10. We find $\mathbb{Z}_2$=1 and $\mathbb{Z}_2$=0 for n_1=0 and n_2=1 respectively, so that the topological class is $1;(000)$. The behavior in Panel (b) is rather uninteresting, since the gap is large everywhere on the n_2=1 face. However, in Panel (a) we again find an example of a rapid change of WCCs with k_2, which was repaired by inserting an extra point (the one now labeled '1' on the horizontal axis) at $k_2 = 0.05\pi$. Actually, we anticipated the need for this denser sampling for small k_2 from the fact that the zero-strain gap closure occurs at Γ, so that a delicate dependence on $\mathbf{k}$ near the BZ center was expected.

Chapter 6

Wannier functions for topological insulators

We now turn to the question of constructing Wannier functions (WFs) for $\mathbb{Z}_2$ topological insulators (TIs). As mentioned in Chapter 4, WFs proved to be an indispensable tool when working with semiconductors and insulators, providing a real-space description of the material that can be used to describe bonding, construct model Hamiltonians and directly compute a number of physical properties, such as polarization [32, 33]. Thus, it is desirable to understand the construction of the Wannier representation of TIs to make them accessible to the plethora of techniques accessible via WFs.

It was discussed in Sec. 4.3 that for Chern insulators a non-zero Chern number presents a topological obstruction that prevents the construction of exponentially localized WFs [37, 38]. Conversely, a general proof has been given that exponentially localized WFs should exist in any 2D or 3D insulator having a vanishing Chern index [36]. In principle this applies to $\mathbb{Z}_2$-odd as well as $\mathbb{Z}_2$-even T-invariant insulators, suggesting that a Wannier representation should be possible in both cases. However, it is unclear whether the nontrivial topology of the $\mathbb{Z}_2$-odd case has any effect on the Wannier representation. In particular, one may wonder whether the procedure for obtaining WFs would be the same as for ordinary insulators, and if not, how it should be modified in order to get well localized WFs in the $\mathbb{Z}_2$-odd regime.

In this chapter we address this question using the model of Kane and Mele [1] (described in Sec. 3.4.3) as a paradigmatic system that exhibits both $\mathbb{Z}_2$-odd and

$\mathbb{Z}_2$-even phases. We demonstrate that the usual projection scheme used for constructing the Wannier representation is still applicable to the $\mathbb{Z}_2$-odd insulators, but only for gauge choices that do not allow WFs to come in time-reversal pairs. We present an explicit projection procedure for constructing well-localized WFs in the topologically non-trivial phase, and show that the WFs can be made even more localized using the standard maximal-localization procedure [32]. We also discuss the electric polarization from both Berry-phase and Wannier points of view, showing the relations between the viewpoints and confirming that both give identical results.

6.1 Explicit construction of Wannier functions

We now consider the problem of constructing WFs starting from a set of Bloch-like states represented on a **k**-mesh. In the approach, which we adopt here, one chooses some localized trial functions in order to provide a starting guess about where the electrons are localized in the unit cell, and obtains a fairly well-localized set of WFs by a projection procedure to be described shortly. If desired, one can follow this with an iterative procedure to make the resulting WFs optimally localized [32] according to the criterion of Sec. 4.1.

Consider an insulator with $\mathcal{N}$ occupied bands. We start with a set of $\mathcal{N}$ trial states $|\tau_i\rangle$ located in the home unit cell, and at each $\mathbf{k}$ we project them onto the occupied subspace at $\mathbf{k}$ to get a set of Bloch-like states

$$|\Upsilon_{i\mathbf{k}}\rangle = \hat{P}_{\mathbf{k}}\,|\tau_i\rangle = \sum_{n=1}^{\mathcal{N}} |\psi_{n\mathbf{k}}\rangle\langle\psi_{n\mathbf{k}}|\tau_i\rangle. \tag{6.1}$$

Since this set of states will not generally be orthonormal, we make use of a Löwdin orthonormalization procedure which consists of constructing the overlap matrix

$$S_{mn}(\mathbf{k}) = \langle\Upsilon_{m\mathbf{k}}|\Upsilon_{n\mathbf{k}}\rangle \tag{6.2}$$

and obtaining the orthonormal set of Bloch-like orbitals

$$|\tilde{\psi}_{n\mathbf{k}}\rangle = \sum_m \left[S(\mathbf{k})^{-1/2}\right]_{mn} |\Upsilon_{m\mathbf{k}}\rangle. \tag{6.3}$$

Note that the $\tilde{\psi}_{n\mathbf{k}}$ are not eigenstates of the Hamiltonian, but they span the same space, and have the same form, as the usual Bloch eigenstates. For an insulator whose gap is not too small, and for a set of trial functions embodying a reasonable assumption about character of the localized electrons, the $\tilde{\psi}_{n\mathbf{k}}$ will be smooth functions of $\mathbf{k}$. In that case, by the usual properties of Fourier transforms, the WFs constructed in analogy with Eq. (4.1),

$$|\mathbf{R}n\rangle = \frac{A}{(2\pi)^2} \int_{BZ} d\mathbf{k}\, e^{-i\mathbf{k}\cdot\mathbf{R}}\, |\tilde{\psi}_{n\mathbf{k}}\rangle, \tag{6.4}$$

should be well localized. Here A is the unit cell area.

Such a construction will break down if the determinant of $S(\mathbf{k})$ vanishes at any $\mathbf{k}$. This is guaranteed to occur in a Chern insulator, where time-reversal symmetry is broken and the Chern index of the occupied manifold is non-zero; in this case, construction of exponentially localized WFs becomes impossible [36–38]. For a $\mathbb{Z}_2$ insulator, however, the presence of time-reversal symmetry guarantees a zero Chern index, so that exponentially localized WFs must exist [36]. In this case, we should be able to find a set of trial functions such that $\det S(\mathbf{k}) \neq 0$ throughout the BZ.

6.1.1 $\mathbb{Z}_2$-even phase

Let us first apply the method described above to the case of the $\mathbb{Z}_2$-even phase of the Kane-Mele (KM) model. This phase is topologically equivalent to the ordinary insulator, so we anticipate a picture in which the two electrons per cell are opposite-spin ones approximately localized on the lower-energy (B) site. One

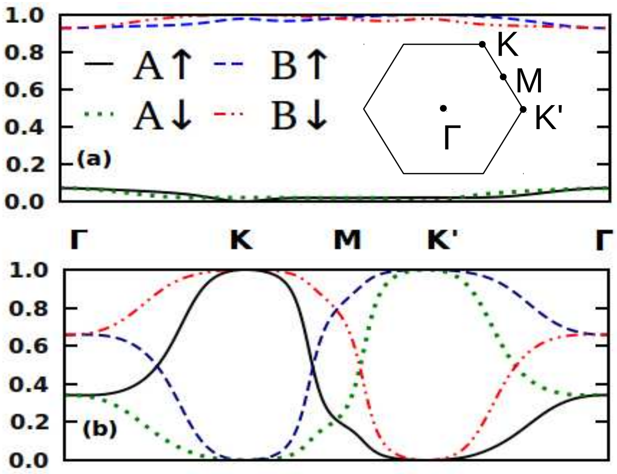

Figure 6.1: Sum of the weights of the projections into the two occupied bands of the basis states $|A;\uparrow_z\rangle$, $|B;\uparrow_z\rangle$, $|A;\downarrow_z\rangle$, and $|B;\downarrow_z\rangle$ plotted along the diagonal of the BZ for (a) $\lambda_v/t = 5$ ($\mathbb{Z}_2$-even phase) and (b) $\lambda_v/t = 1$ ($\mathbb{Z}_2$-odd phase). Inset in (a): BZ of a honeycomb lattice.

way to see this is to look at the weights of the basis states in the occupied subspace. Figure 6.1(a) shows the distribution of these weights along a high-symmetry line in the BZ for the KM model in its $\mathbb{Z}_2$-even phase. From the figure it is obvious that the two basis states on the B site dominate in the occupied subspace over the whole BZ. It is then natural to choose the two trial functions to be opposite-spin spatial δ-functions localized on the B site in the home unit cell. We choose these to be spin-aligned along z, i.e.,

$$|\tau_i\rangle = |B;\sigma_i^z\rangle = \delta(\mathbf{r}-\mathbf{t}_B)|\sigma_i^z\rangle \tag{6.5}$$

where $|\sigma_1^z\rangle = |\uparrow_z\rangle$ and $|\sigma_2^z\rangle = |\downarrow_z\rangle$. Transforming to $\mathbf{k}$-space we get

$$|\tau_{i\mathbf{k}}\rangle = \frac{|\sigma_i^z\rangle}{\sqrt{N}}\sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}}\delta(\mathbf{r}-\mathbf{R}-\mathbf{t}_B). \tag{6.6}$$

The two occupied Bloch bands may be written as

$$|\psi_{n\mathbf{k}}\rangle = \sum_{\ell} C_{\ell n\mathbf{k}} |\chi_{\ell\mathbf{k}}\rangle \tag{6.7}$$

where ℓ is a combined index for sublattice and spin, $\ell = \{1, 2, 3, 4\} \equiv \{A\uparrow, B\uparrow, A\downarrow, B\downarrow\}$, and $\chi_{\ell\mathbf{k}} = \chi_{j\sigma\mathbf{k}}$ are the tight-binding basis functions of Eq. (3.33).[1] With Eq. (6.6) the projected functions become

$$|\Upsilon_{1\mathbf{k}}\rangle = C^*_{21\mathbf{k}}|\psi_{1\mathbf{k}}\rangle + C^*_{22\mathbf{k}}|\psi_{2\mathbf{k}}\rangle, \tag{6.8}$$

$$|\Upsilon_{2\mathbf{k}}\rangle = C^*_{41\mathbf{k}}|\psi_{1\mathbf{k}}\rangle + C^*_{42\mathbf{k}}|\psi_{2\mathbf{k}}\rangle. \tag{6.9}$$

The overlap matrix S is constructed from these functions, and for the determinant one finds

$$\begin{aligned} \det\left[S(\mathbf{k})\right] &= (|C_{21\mathbf{k}}|^2 + |C_{22\mathbf{k}}|^2)(|C_{41\mathbf{k}}|^2 + |C_{42\mathbf{k}}|^2) \\ &- |C_{21\mathbf{k}}C^*_{41\mathbf{k}} + C_{22\mathbf{k}}C^*_{42\mathbf{k}}|^2. \end{aligned} \tag{6.10}$$

Recall that for the Löwdin orthonormalization procedure to succeed, this determinant must remain non-zero everywhere in the BZ. This is indeed the case for the $\mathbb{Z}_2$-even phase, as illustrated in Fig. 6.2(a), where the solid black curve shows the dependence of the determinant on $\mathbf{k}$ along the high-symmetry line in the BZ.

In contrast, the dashed red curve in Fig. 6.2(a) shows the behavior of $\det\left[S(\mathbf{k})\right]$ in the $\mathbb{Z}_2$-odd regime. The determinant can be seen to vanish at the K and K' points in the BZ. Clearly, this choice of trial functions is not appropriate for building the Wannier representation in the $\mathbb{Z}_2$-odd phase. Indeed, as we shall see in the next section, *any* choice of trial functions that come in Kramers pairs is guaranteed to fail in the $\mathbb{Z}_2$-odd case. There we shall also investigate alternative

[1] In the sense of Sec. 2.3 there are two orbitals per cite: one with spin up, another with spin down.

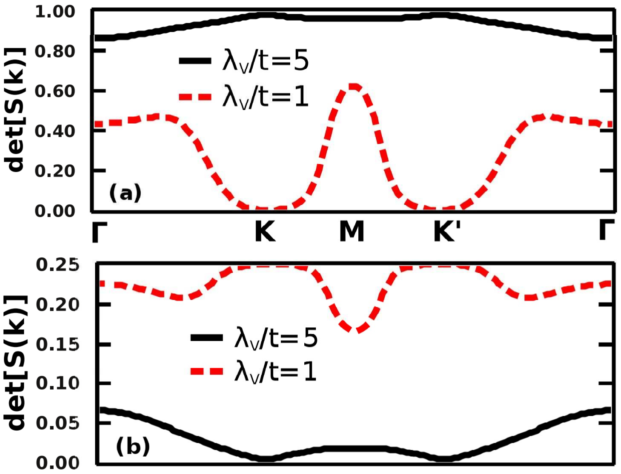

Figure 6.2: Plot of det$[S(\mathbf{k})]$ along the diagonal of the BZ for $\lambda_v/t = 5$ ($\mathbb{Z}_2$-even phase) and $\lambda_v/t = 1$ ($\mathbb{Z}_2$-odd phase). (a) Trial functions are $|B;\uparrow_z\rangle$ and $|B;\downarrow_z\rangle$. (b) Trial functions are $|A;\uparrow_x\rangle$ and $|B;\downarrow_x\rangle$.

choices of trial functions that allow for a successful construction of WFs.

6.1.2 $\mathbb{Z}_2$-odd phase

To gain some insight into the appropriate choice of trial functions in the $\mathbb{Z}_2$-odd regime, consider the weights of the basis functions in the occupied space shown for this case in Fig. 6.1(b). Unlike the normal insulator, the $\mathbb{Z}_2$-odd phase does not favor any particular basis states. Instead, different basis states dominate in different portions of the BZ. For example, at points K and K' the occupied space is represented by only two of the four basis states; at each of these points the two participating basis states have opposite spin and sublattice indices, and none appear in common at both points. (The states at K are, of course, Kramers pairs of those at K'.) It follows that if any of the trial states is simply set equal to one of the four basis states, then at least one of the $|\Upsilon\rangle$ would vanish either at K or K', and the determinant would vanish there too. This explains the failure of the naive Wannier construction procedure for the $\mathbb{Z}_2$-odd phase; with the naive choice of trial functions as in Eq. (6.5), the determinant vanishes at both K and

K', as shown by the red dashed curve in Fig. 6.2(a).[2]

In fact, this failure can be understood from a general point of view. If the two trial functions form a Kramers pair, then the projection procedure of Eqs. (6.1-6.3) will result in Bloch-like functions obeying

$$\begin{aligned} |\tilde{\psi}_1(-\mathbf{k})\rangle &= \theta|\tilde{\psi}_2(\mathbf{k})\rangle, \\ |\tilde{\psi}_2(-\mathbf{k})\rangle &= -\theta|\tilde{\psi}_1(\mathbf{k})\rangle. \end{aligned} \tag{6.11}$$

The WFs obtained from Eq. (6.4) will then also form a Kramers pair. But Eq. (6.11) is nothing other than the constraint of Eq. (3.39) defining a gauge that respects time-reversal symmetry, and it has been shown [27, 110, 116] that an odd value of the $\mathbb{Z}_2$ invariant presents an obstruction against constructing such a gauge. In other words, in the $\mathbb{Z}_2$-odd phase a smooth gauge cannot be fixed by choosing trial functions that are time-reversal pairs of each other, and a choice of WFs as time-reversal pairs is not possible. Hence, in order to construct the Wannier representation in the $\mathbb{Z}_2$-odd regime, one should choose trial functions that do not transform into one another under time reversal.

Following these arguments, we choose the two trial functions to be localized on *different* sites in the home unit cell. Moreover, in order that they will have components on states with spins both up and down along z, we choose the spins of the trial states so that one is along $+x$ and the other along $-x$.[3] In $\mathbf{k}$-space this becomes

$$|\tau_{i\mathbf{k}}\rangle = \frac{|\sigma_i^x\rangle}{\sqrt{N}} \sum_{\mathbf{R}} e^{i\mathbf{k}\cdot\mathbf{R}} \delta(\mathbf{r} - \mathbf{R} - \mathbf{t}_i) \tag{6.12}$$

[2] Fig. 6.1(b) shows also that the character of the occupied states changes in the BZ, which serves an illustration of band inversion, associated with TIs.

[3] To see that the z components have to be mixed, consider two trial functions that are localized on different sites A and B with opposite direction of spin in the z direction. But in this case the projected functions $\Upsilon_{i\mathbf{k}}$ become zero either at K or K', which follows from the Fig. 6.1(b).

where $\mathbf{t}_1 = \mathbf{t}_A$ and $\mathbf{t}_2 = \mathbf{t}_B$, leading to

$$|\Upsilon_{1\mathbf{k}}\rangle = \left[(C^*_{11\mathbf{k}} + C^*_{31\mathbf{k}})|\psi_1\rangle + (C^*_{12\mathbf{k}} + C^*_{32\mathbf{k}})|\psi_2\rangle\right]/\sqrt{2} \tag{6.13}$$

and

$$|\Upsilon_{2\mathbf{k}}\rangle = \left[(C^*_{21\mathbf{k}} - C^*_{41\mathbf{k}})|\psi_1\rangle + (C^*_{22\mathbf{k}} - C^*_{42\mathbf{k}})|\psi_2\rangle\right]/\sqrt{2}. \tag{6.14}$$

The determinant takes the form

$$\begin{aligned} \det[S] &= (|C_{11\mathbf{k}} + C_{31\mathbf{k}}|^2 + |C_{12\mathbf{k}} + C_{32\mathbf{k}}|^2)(|C_{21\mathbf{k}} - C_{41\mathbf{k}}|^2 + |C_{22\mathbf{k}} - C_{42\mathbf{k}}|^2)/4 - \\ &- |(C_{11\mathbf{k}} + C_{31\mathbf{k}})(C^*_{21\mathbf{k}} - C^*_{41\mathbf{k}}) + (C_{12\mathbf{k}} + C_{32\mathbf{k}})(C^*_{22\mathbf{k}} - C^*_{42\mathbf{k}})|^2/4. \quad (6.15) \end{aligned}$$

The dependence $\det[S(\mathbf{k})]$ is shown along the diagonal of the Brillouin zone for this choice of trial functions in Fig. 6.2(b). In the $\mathbb{Z}_2$-odd phase (dashed line) the determinant remains non-zero everywhere in the BZ.[4] Not surprisingly, the same trial functions are very poorly suited to the normal-insulator phase, as can be seen from solid line in the same panel. In this case $\det[S(\mathbf{k})]$ almost vanishes at K and K' and remains quite small throughout the rest of the BZ, so that one should clearly revert to the time-reversed pair of trial functions of Eq. (6.5) and Fig. 6.2(a) in order to get well-localized WFs.

We made an arbitrary choice above in selecting the two trial functions to be up and down along x. In fact, if we repeat the entire procedure using trial functions that are spin-up and spin-down along any unit vector $\hat{n}$ lying in the xy-plane, we find that $\det[S(\mathbf{k})]$ changes very little, with only small changes in the size of the dip near the Γ point. Thus, we find that the choice of trial functions in Eq. (6.12) is not unique. Instead, there is a large degree of arbitrariness in the choice of WFs in the $\mathbb{Z}_2$-odd case.

To conclude, we have established that the choice of a time-reversal pair of

[4] Actually, the minimum value of $|\det[S]|$ in the BZ is equal to 0.0873.

trial functions, Eq. (6.5), that allows for the construction of well-localized WFs in the ordinary-insulator phase cannot be used in the $\mathbb{Z}_2$-odd phase. In order for the usual projection method for constructing the Wannier representation to work in this topologically nontrivial phase, the trial functions should explicitly break time-reversal symmetry, i.e., they should not come in time-reversal pairs.

6.2 Localization of Wannier Functions in the $\mathbb{Z}_2$-odd Insulator

Now that we know how to construct WFs for the $\mathbb{Z}_2$-odd insulator, we discuss their localization properties. As we have noted in the preceding section, the choice of the trial functions, Eq. (6.12), is not unique; there are other gauge choices arising from different trial functions that also produce well-defined sets of WFs. Since different gauge choices lead to different degrees of localization of the resulting WFs, it is natural to fix the gauge by the condition of maximal possible localization of the WFs.

The problem of constructing maximally-localized WFs was discussed in Sec. 4.1. It requires the iterative minimization of the spread functional 4.3 with respect to a U($\mathcal{N}$) transformations of the occupied space. In order to avoid getting trapped in false local minima, the iterative procedure is normally initialized using the trial-function projection procedure described in Sec. 6.1 above.

We now apply this method to the KM model. The lattice is hexagonal, and in this case six $\mathbf{b}_j$ vectors are needed to satisfy the condition (4.8), namely $\mathbf{b}_1 = -\mathbf{b}_4 = \mathbf{G}_1/q$, $\mathbf{b}_2 = -\mathbf{b}_5 = (\mathbf{G}_1 + \mathbf{G}_2)/q$, and $\mathbf{b}_3 = -\mathbf{b}_6 = \mathbf{G}_2/q$. All six have the same length b and weight $\omega_b = 1/(3b^2)$. We start with the WFs obtained with the projection method using the trial functions of Eq. (6.12), appropriate for the $\mathbb{Z}_2$-odd phase.

The resulting spreads, both before and after the iterative minimization, are

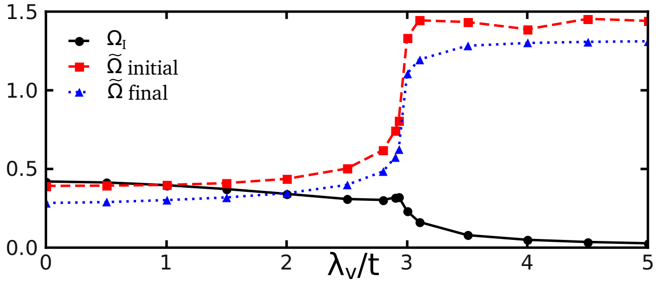

Figure 6.3: Wannier spreads Ω_I and $\tilde{\Omega}$ for the Kane-Mele model on a 60×60 **k**-mesh, initialized using the trial functions of Eq. (6.12). "Initial" and "final" values are those computed before and after the iterative minimization respectively. The system is in the $\mathbb{Z}_2$-odd phase for $\lambda_v/t \lesssim 2.93$.

shown in Fig. 6.3. (Ω_I, being gauge-invariant, is the same before and after.) The left part of the figure shows the behavior in the $\mathbb{Z}_2$-odd phase, where the trail functions are the appropriate ones. The results in this region were not strongly sensitive to the **k**-point mesh density. The fact that $\tilde{\Omega}$ is similar in magnitude to the unminimized Ω_I, and that the localization procedure reduces $\tilde{\Omega}$ by only $20-30\%$, provide additional evidence that the choice of trial functions was a good one. The Wannier charge centers were almost unchanged by the minimization procedure; the x-coordinates were zero, while $\bar{r}_{1y} \simeq a/\sqrt{3}$ and $\bar{r}_{2y} \simeq 2a/\sqrt{3}$ (see Sec. 6.3 for details), in good agreement with our initial assumption about the WFs being localized on A and B sites.

The right part of Fig. 6.3, for $\lambda_v/t \gtrsim 2.93$, shows what happens when we attempt to use the same trial functions in the normal phase. Ω_I is of course unaffected by the choice of trial functions, and the fact that it has a smaller value in this region indicates, not surprisingly, that the insulating state is simpler and more localized in the normal state. (For large λ_v/t the WFs approach spatial delta functions, explaining the fact that Ω_I asymptotes to zero in that limit.) Not surprisingly, however, using the trial functions appropriate to the $\mathbb{Z}_2$-odd phase in the $\mathbb{Z}_2$-even regime results in very poor localization of the WFs as measured by $\tilde{\Omega}$. Our data also suggests that in the $\mathbb{Z}_2$-odd phase MLWFs are less localized than MLWS in the $\mathbb{Z}_2$-even phase. For example, the use of trial functions (6.5) with $\lambda_v/t = 5$ and a 60×60 $\mathbf{k}$-mesh results in $\Omega_I = 0.027695$ and $\tilde{\Omega} = 0.000249$. We also find that the results are more sensitive to the choice of $\mathbf{k}$-mesh in the $\mathbb{Z}_2$-odd regime.

To summarize the results of this section, we studied the construction of maximally localized WFs in the $\mathbb{Z}_2$-odd phase using the KM model as an example. We have seen that our initial guess of Sec. 6.1 about the localization of WFs in this topological regime is very good, and that the maximal localization procedure does not greatly reduce the spread.

6.3 Hybrid WCCs and polarization of Kane-Mele model

In this section we discuss the polarization in $\mathbb{Z}_2$-odd insulators using the example of the KM model, and see what insights about the topological insulating phase can be obtained by inspecting this property.

The electronic polarization in a 2D system can be defined either in terms of the Berry phase [10]

$$\mathbf{P} = \frac{|e|}{(2\pi)^2}\mathrm{Im}\sum_{n=1}^{\mathcal{N}}\int d\mathbf{k}\langle u_{n\mathbf{k}}|\nabla_{\mathbf{k}}|u_{n\mathbf{k}}\rangle \tag{6.16}$$

or via the summation of Wannier charge centers [33]

$$\mathbf{P} = -\frac{|e|}{A}\sum_{n=1}^{\mathcal{N}}\bar{\mathbf{r}}_n, \tag{6.17}$$

where e is the electronic charge and A is the area of the unit cell. The two definitions are identical and define electronic polarization modulo a polarization quantum $|e|\mathbf{R}/A$, $\mathbf{R}$ being a lattice vector. This ambiguity can be understood as a freedom in the choice of branch in Eq. (6.16) or in the choice of unit cell in

Eq. (6.17). The definition via Wannier charge centers makes the dependence of $\mathbf{P}$ on the choice of origin obvious. As described in Sec. 3.4.3, the origin of the KM model is chosen such that atoms are located along the y-axis at $\mathbf{t}_A = \xi\hat{y}/3$ and $\mathbf{t}_B = 2\xi\hat{y}/3$, where $\xi = |\mathbf{a}_1 + \mathbf{a}_2| = a\sqrt{3}$ (see Fig. 3.1). Because the Hamiltonian has 3-fold symmetry, we expect the rescaled polarization $(A/|e|)\mathbf{P}$ to lie at the origin, at $\mathbf{t}_A$, or at $\mathbf{t}_B$. To distinguish between these possibilities it is sufficient to compute P_y, which is well-defined modulo $|e|/a$.

6.3.1 Total polarization

A direct computation of electronic polarization via Eq. (6.16) in the $\mathbb{Z}_2$-even phase results in $P_y = |e|/3a \bmod |e|/a$, consistent with the fact that both Wannier centers in Eq. (6.17) lie at $\mathbf{t}_B$ (since $-4|e|\xi/3A = -8|e|/3a = |e|/3a \bmod |e|/a$.) In the $\mathbb{Z}_2$-odd phase, on the other hand, Eqs. (6.16) and (6.17) lead to $P_y = 0 \bmod |e|/a$. Again, this is consistent with the locations of the WFs. As indicated in Sec. 6.2, the Wannier centers $\bar{\mathbf{r}}_n$ in this phase lie approximately at $\mathbf{t}_A$ and $\mathbf{t}_B$. More precisely, we find that they are located at $\bar{\mathbf{r}}_1 = (1-\delta)\xi\hat{y}/3$ and $\bar{\mathbf{r}}_2 = (2+\delta)\xi\hat{y}/3$, where δ is a small correction (e.g., $\delta = 0.0018$ at $\lambda_v/t = 1$). Thus, the sum of the Wannier centers is just $\xi\hat{y}$, or zero modulo a lattice vector.

It is interesting to note that, in retrospect, the computation of the polarization via Eq. (6.16) would have given a strong hint about the appropriate choice of trial functions in the $\mathbb{Z}_2$-odd insulator. That is, knowing only that $P_y = 0$, one might have guessed that both WFs should be centered halfway between $\mathbf{t}_A$ and $\mathbf{t}_B$, or both at the center of the honeycomb ring, or one at $\mathbf{t}_A$ and the other at $\mathbf{t}_B$. The latter possibility becomes the most likely when we also take into account that in the $\mathbb{Z}_2$-odd phase the two WFs cannot form a Kramers pair.

6.3.2 Hybrid Wannier decomposition

In order to obtain a deeper understanding of the origin of the polarization and expose some qualitative differences in the behavior of its $\mathbf{k}$-dependent decomposition in $\mathbb{Z}_2$-even and odd phases, it is useful to use a hybrid representation in which the Wannier transformation is carried out in one direction only. As indicated above, we know from symmetry considerations that we can set $P_x = 0$ and characterize the polarization by P_y mod $\xi|e|/A$. To compute P_y, it is convenient to choose the BZ to be a rectangle extending over $k_x \in [0, 2\pi/a]$ and $k_y \in [0, 4\pi/\xi]$ (corresponding to the region ζ in Fig. 3.6). In our notations the hybrid WFs defined by Eq. 4.10

$$|R_y k_x n\rangle = \frac{\xi}{4\pi} \int_0^{4\pi/\xi} dk_y \, e^{-ik_y R_y} |\tilde{\psi}_{n\mathbf{k}}\rangle. \tag{6.18}$$

The hybrid Wannier centers are defined as

$$\bar{y}_n(k_x) = \langle 0k_x n|y|0k_x n\rangle \tag{6.19}$$

and the total electronic polarization is

$$P_y = -\frac{|e|}{\pi\xi} \sum_n \int_0^{2\pi/a} dk_x \, \bar{y}_n(k_x). \tag{6.20}$$

In practice the k_x integral is discretized by a sum over a mesh of k_x values, and at each k_x the $\bar{y}_n(k_x)$ are calculated by considering the corresponding string of $\mathbf{k}$-points along k_y. In the case that the gauge has been specified by a particular set of 2D WFs $|\mathbf{R}n\rangle$, or, equivalently, by the corresponding Bloch-like functions $|\tilde{\psi}_{n\mathbf{k}}\rangle$, this is done straightforwardly using the discretized Berry-phase formula

$$\bar{y}_n(k_x) = -\frac{\xi}{4\pi} \operatorname{Im} \log \prod_j M_{nn}^{(j)} \tag{6.21}$$

where $M^{(j)}$ is a shorthand for the overlap matrix $M^{(\mathbf{k}_j,\mathbf{k}_{j+1})}$ of Eq. (4.7) connecting k_y-points j and $j+1$ along the string.

As was emphasized in Sec. 6.1, the $\tilde{\psi}_{n\mathbf{k}}$ carry the information about the gauge choice. Thus, different gauge choices – i.e., different choices of WFs – will result in different hybrid WFs and different $\bar{y}_n(k_x)$. However, the sum $\sum_n \bar{y}_n(k_x)$ at a given k_x is gauge-invariant, and as a result P_y of Eq. (6.20) must remain the same in any gauge.

Of special interest is a gauge choice in which, at each k_x, the hybrid WFs $|nk_xl_y\rangle$ are maximally localized in the y direction. This is the parallel transport gauge of Sec. 4.4, where the states are transported along the k_y direction. The Wannier charge centers are defined [117] by the eigenvalues λ_n of the matrix Λ from Eq. (4.23) via

$$\bar{y}_n(k_x) = -\frac{\xi}{4\pi}\text{Im}\log\lambda_n(k_x). \tag{6.22}$$

Note that no iterative procedure is needed. Inserting this equation into Eq. (6.20), one gets a discretized formula for P_y that is consistent with Eq. (6.16).

6.3.3 Results

We illustrate these ideas now for the KM model in its normal and $\mathbb{Z}_2$-odd phases. In each case we present results for $\bar{y}_n(k_x)$ for two choices of gauge: the maximally-localized one along $\hat{y}$ as discussed in the previous paragraph, and the one corresponding to the WFs constructed from the trial functions of Eq. (6.5) for the $\mathbb{Z}_2$-even phase or those of Eq. (6.12) for the $\mathbb{Z}_2$-odd phase. In what follows, we refer to these as the “maxloc” and “WF-based” gauges respectively.

In the ordinary insulating regime, the maxloc and WF-based $\bar{y}_n(k_x)$ curves look very similar to each other. Fig. 6.4(a) and (b) show the calculated results for the case of $\lambda_v/t = 3$, very close to the transition on the insulating side (recall the critical value is at $\lambda_v/t = 2.93$). Three of the infinite number of periodic

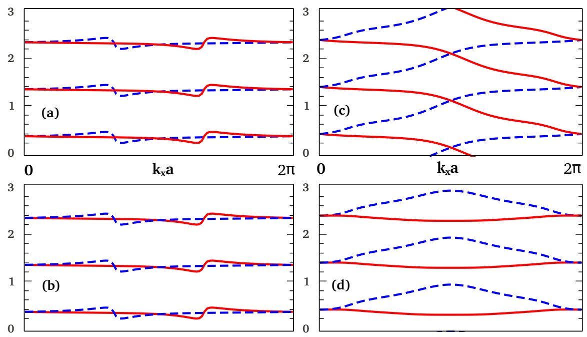

Figure 6.4: Hybrid Wannier centers $\bar{y}_n(k_x)$, in units of $\xi/2$, for the Kane-Mele model. $\mathbb{Z}_2$-even phase ($\lambda_v/t = 3$): (a) maxloc gauge; (b) WF gauge of Eq. (6.5). $\mathbb{Z}_2$-odd phase ($\lambda_v/t = 1$): (c) maxloc gauge; (d) WF gauge of Eq. (6.12). In each case, several periodic images are shown.

images along y are shown. The "bumps" in the curves near the K and K' points in the BZ are the result of the proximity to the transition; as one goes deeper into the insulating phase, the curves flatten out and become smooth functions of k_x. The solid and dashed curves are mirror images of each other; in the maxloc construction of Fig. 6.4(a) this just reflects the time-reversal invariance of the Hamiltonian, while in Fig. 6.4(b) it follows from the fact that the WFs form a Kramers pair.

When averaged over k_x, each curve is found to have a mean $\bar{y}$ value of $2\xi/3$ to numerical precision, or $\xi/6$ modulo $\xi/2$, consistent with the discussion in Sec. 6.3.1.

The corresponding results for the $\mathbb{Z}_2$-odd phase are shown in Fig. 6.4(c) and (d) for $\lambda_v/t = 1$. As expected, there is again a mirror symmetry visible in the curves for the maxloc construction in Fig. 6.4(c), but the connectivity of the curves is qualitatively different: in going from $k_x = 0$ to π/a we see that the n'th solid curve goes up to cross the $(n+1)$'th dashed curve, while the n'th dashed curve goes down to cross the $(n-1)$'th solid curve. This is exactly the kind of behavior that was exhibited in Fig. 3(a) of Ref. [27] and discussed in Chapter 5 as a signal of the $\mathbb{Z}_2$-odd phase. Moreover, if we follow the n'th dashed curve all the way across the BZ, we find that it wraps to become the $(n+1)$'th one when $k_x = 2\pi/a$ wraps back to $k_x = 0$. This is precisely the kind of behavior that is characteristic of a Chern (or quantum anomalous Hall) insulator [106], which implies that we can assign a Chern number of $+1$ to the Bloch subspace spanned by the eigenvectors corresponding to the dashed bands. However, since we are studying here a system with time-reversal symmetry, we find also a partner subspace corresponding to the full curve in Fig. 6.4(c) having Chern number -1. As a result, of course, the overall occupied space has a total vanishing Chern number, as it must due to the time-reversal symmetry. The evaluation of the polarization P_y through Eq. (6.20) again yields $P_y = 0 \bmod |e|/a$, consistent with the direct calculation of Sec. 6.3.1.

Finally, Fig. 6.4(d) shows the $\bar{y}_n(k_x)$ curves for the same $\mathbb{Z}_2$-odd parameters as in Fig. 6.4(c), but using the WF-based gauge determined by the trial functions of Eq. (6.12). At any given k_x, we confirm that $\bar{y}_1 + \bar{y}_2$ is the same in Fig. 6.4(d) as in Fig. 6.4(c), and the total polarization is therefore the same. However, because the two WFs do not form a Kramers pair in this case, the dashed and solid curves do not map into each other under time-reversal symmetry, and there is no degeneracy at $k_x = \pi/a$. Moreover, the Chern number of each band is individually zero, consistent with the fact that each one is derived from a WF. This illustrates the point made in Chapter 5 that the $\mathcal{T}$ symmetric gauge is crucial for the $\mathcal{T}$ polarization approach to the $\mathbb{Z}_2$ invariant. The average $\bar{y}$ values for the solid and dashed curves are $0.978\xi/3$ and $2.022\xi/3 \bmod \xi/2$, very close to the nominal locations of the trial functions at $\mathbf{t}_A$ and $\mathbf{t}_B$, respectively.

To recap, in both the $\mathbb{Z}_2$-even and $\mathbb{Z}_2$-odd cases, we find that the occupied Bloch space can be cast as the direct sum of two subspaces that map into one another under the time-reversal operation, corresponding to the solid and dashed curves of Figs. 6.4(a-c). These subspaces are not built from Hamiltonian eigenstates, but from suitable $\mathbf{k}$-dependent U(2) rotations among the Hamiltonian eigenstates. In the $\mathbb{Z}_2$-even case the Chern index of each of these subspaces is separately zero, so that we can also provide a Wannier representation for each subspace separately. This is essentially the case of Fig. 6.4(b), and since the spaces form a time-reversal pair, the WFs form a time-reversal pair as well. In contrast, for the $\mathbb{Z}_2$-odd phase, the decomposition into two subspaces that are time-reversal images of each other necessarily results in subspaces having individual Chern numbers of ± 1, and these are not individually Wannier-representable. Only by violating the condition that the two spaces be time-reversal partners, as was done in Fig. 6.4(d), can we decompose the space into two subspaces having zero Chern indices individually. By doing so, we can find a Wannier representation of the entire space, but only on condition that the two WFs do not form a Kramers pair.

6.4 Generalization to 3D

The generalization of our findings to the 3D case should be relatively straightforward. Certainly the topological obstruction to the construction of Kramers-pair WFs remains for both weak and strong $\mathbb{Z}_2$ TIs in 3D. To see this, consider in turn each of the six symmetry planes in $\mathbf{k}$-space ($k_1 = 0$, $k_2 = 0$, $k_3 = 0$, $k_1 = \pi/a$, etc.) on which $H_{\mathbf{k}}$ behaves like a 2D time-reversal invariant system. For both weak and strong TIs, at least one of these six planes must have a $\mathbb{Z}_2$-odd 2D invariant. But if a gauge exists obeying the time-reversal condition of Eq. (3.39) in the 3D $\mathbf{k}$-space, then it does so in particular on the 2D plane, in contradiction with the 2D arguments about a topological construction.

Thus, the general strategy for constructing WFs for 3D topological insulators should be very similar to the one presented here in 2D. Namely, one has to choose pairs of trial functions that do not transform into one another by time-reversal symmetry, and to do it in such a way that the projection of these trial functions onto the Bloch states does not become singular anywhere in the 3D BZ. While it may be interesting to explore how this might best be done in practice for real 3D TIs, e.g., in the density-functional context, the choice is likely to depend sensitively on details of the particular system of interest.

Chapter 7
Smooth gauge for topological insulators

For an ordinary insulator the Bloch states $\psi_{n\mathbf{k}}$ are usually assumed to be smooth and periodic in the Brillouin zone (BZ), meaning that a translation by a reciprocal lattice vector $\mathbf{G}$ returns the Bloch wavefunction back to itself with the same phase, $\psi_{n,\mathbf{k}+\mathbf{G}} = \psi_{n\mathbf{k}}$, and that ψ is a smooth function of $\mathbf{k}$. Regarding the BZ as a torus, as in Fig. 7.1(a), this just means that ψ is a smooth function of $\mathbf{k}$ on the torus. This turns out to be impossible for Chern insulators [37, 38]; the occupied space of a Chern insulator cannot be represented by smooth and periodic Bloch states. Usually periodicity is still assumed, in which case a point discontinuity or branch cut must appear in the phase of at least one occupied Bloch state somewhere in the BZ. It is now established that no gauge transformation – i.e., no $\mathbf{k}$-dependent unitary rotation of the bands in the occupied subspace – can smooth out this discontinuity [37, 38].

In the case of $\mathbb{Z}_2$ insulators, the presence of $\mathcal{T}$ symmetry forces the total

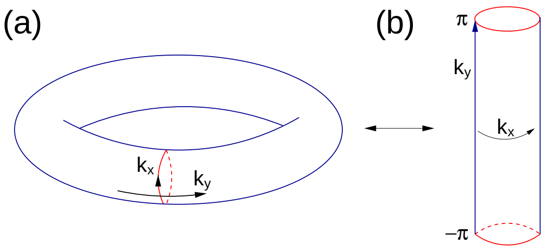

Figure 7.1: Brillouin zone in 2D represented as (a) a torus, and (b) a cylinder. We choose the gauge discontinuity to be distributed along the cross-sectional cut of the torus that maps onto the end loops of the cylinder at $k_y = \pm\pi$.

Chern number to vanish, guaranteeing the existence, in principle, of a smooth and periodic gauge in the BZ [36]. However, it has been shown that any gauge that respects $\mathcal{T}$ symmetry cannot be smooth on the torus for this class of topological materials [27, 40, 110, 116]. Thus, the construction has to break $\mathcal{T}$ symmetry if it is to lead to a smooth gauge. An explicit construction of this type for the model of Kane and Mele, presented in the previous chapter, demonstrated that this is possible [40], but the method used there was explicitly model-dependent, and it remained unclear how one should choose a smooth gauge for a generic $\mathbb{Z}_2$ insulator.

In the present chapter we address this question and develop a general procedure for constructing smooth and periodic Bloch states for the quantum spin Hall (QSH) insulators. We limit ourselves to the minimal case of two occupied bands and show how they can be disentangled into two single-band subspaces having equal and opposite Chern numbers, in such a way that these subspaces are mapped onto each other by the $\mathcal{T}$ operator θ.

Each of these "Chern bands" has the same type of gauge discontinuity on the boundary as is present in a Chern insulator. The possibility of such a decomposition has been discussed before in different contexts [81, 110, 118, 119], but the previous approaches all have relied on some specific feature of the system, such as separation of states according to the action of the s_z or mirror symmetry operators. Instead, our construction is based on topological considerations alone, and should remain robust for any $\mathbb{Z}_2$ insulator. We further impose on these Chern bands a special "cylindrical gauge" in which the gauge discontinuity is spread uniformly around the circular cross section of the BZ torus, i.e., connecting the end loops at $k_y = \pm\pi$ in Fig. 7.1(b). Finally, we develop a procedure that mixes these two topologically non-trivial states in such a way that they become smooth and periodic in the BZ, thus obtaining a smooth (but $\mathcal{T}$-broken) gauge.

Apart from the purely theoretical motivation, the problem of constructing smooth Bloch states for $\mathbb{Z}_2$ insulators has a direct practical application. In Chapter 4 we discussed the use of Wannier functions (WFs) for computing many properties of insulating materials [7, 32–34]. However, it was stressed that exponentially localized Wannier functions may be constructed only out of a set of smooth Bloch states. Thus, construction of a smooth gauge for $\mathbb{Z}_2$ insulators allows for the use of well-established Wannier-based methods in the study of these materials. A particular construction of exponentially localized WFs for a QSH phase, presented in Chapter 6, has a significant drawback of being model-dependent. Although, in some cases symmetries of the system might allow one to quickly find suitable initial projections, and construct WFs for the model, it is preferable to have a model independent procedure.

Another interesting aspect of the present development arises from the fact that a smooth gauge allows one to compute the $\mathbb{Z}_2$ topological invariant directly by means of the formula 3.37, revealing the connectivity of states between $\mathcal{T}$ invariant points in the BZ [27]. So far, this direct computation was done only in the presence of inversion symmetry [31], since its presence allows one to choose states that are smoothly connected in the BZ. In the absence of inversion symmetry, however, the same is not true, and the computation of the topological invariant also becomes more complicated [39, 84, 120, 121]. Thus, one can consider the present method as an alternative recipe for computing topological invariants.

The present discussion treats the two-dimensional case. For a three-dimensional $\mathcal{T}$-invariant insulator, the method described here can be used to construct a smooth gauge on any of the six $\mathcal{T}$-invariant planes in the BZ. However, the final connection between these faces to obtain a globally smooth gauge in 3D appears to be nontrivial except in special cases (e.g., certain kinds of weak topological insulators). A general formulation in 3D is therefore left to future investigations.

We repeatedly emphasize the statement of Sec. 3.1 that questions of gauge choice do not affect physical observables such as the dispersions or spin textures of the energy bands. Thus, if used properly, even a gauge that violates $\mathcal{T}$ symmetry, or that has a gauge discontinuity on the BZ boundary, should be capable of making robust predictions of physical properties consistent with $\mathcal{T}$ symmetry. We are concerned here with formal issues of gauge construction and practical questions about which construction is most convenient for computing physical properties.

7.1 Cylindrical gauge and individual Chern numbers

In this section we return to the definition of the Chern number of a Bloch band in 2D, given in Sec. 3.3, and introduce a cylindrical gauge for Chern bands. This is a gauge that is continuous in the BZ but is periodic in k_x only. That is, it is continuous on the cylinder in Fig. 7.1(b), but not across the boundary connecting top to bottom, i.e., not on the torus of Fig. 7.1(a). We then establish the notion of individual band Chern numbers in the multiband case.

7.1.1 Single band case

Let us first consider a single isolated Bloch band $\psi_{\mathbf{k}}(\mathbf{r})$ in 2D and its cell periodic part $u_{n\mathbf{k}}(\mathbf{r}) = e^{-i\mathbf{k}\cdot\mathbf{r}}\psi_{\mathbf{k}}(\mathbf{r})$. We assume the lattice vectors to have unit length and to be aligned with the Cartesian axes, i.e., $\mathbf{a}_1 = \hat{x}$ and $\mathbf{a}_2 = \hat{y}$, so that k_x runs from 0 to 2π and k_y runs from $-\pi$ to π. (In the general case, a linear transformation trivially rescales the k indices into this form.)

For a single band a Chern number of Eq. (3.18) is given by the integral of the Abelian curvature (3.15) over the BZ, represented by a torus in Fig. 7.1(a):

$$C = \frac{1}{2\pi}\int_{BZ} d^2k\mathcal{F}(\mathbf{k}). \tag{7.1}$$

In general, the non-zero Chern number reflects the impossibility of constructing a periodic gauge without the presence of points or lines in the BZ where the wavefunction would have a phase discontinuity.

To have a particular example of a gauge that leads to a nonzero Chern number C, consider a gauge that is smooth everywhere on the BZ torus except on a circle as shown in Fig. 7.1(a). Any gauge discontinuity that might be present has thus been pushed to this circular boundary, where the phase of the wavefunction can experience a jump when crossing it. Such a gauge is continuous on the cylinder formed by cutting the torus along the discontinuity, shown in Fig. 7.1(b), but is not periodic in the y direction. We now define a "cylindrical" gauge to be one in which the gauge discontinuity is uniformly distributed around the boundary. That is, such a gauge obeys the boundary conditions

$$\begin{aligned} \psi_{\mathbf{k}+2\pi\hat{x}} &= \psi_{\mathbf{k}}, \\ \psi_{\mathbf{k}+2\pi\hat{y}} &= \psi_{\mathbf{k}}\, e^{iCk_x}, \end{aligned} \tag{7.2}$$

or, equivalently,

$$\begin{aligned} u_{\mathbf{k}+2\pi\hat{x}} &= e^{-2\pi ix}\, u_{\mathbf{k}}, \\ u_{\mathbf{k}+2\pi\hat{y}} &= e^{-2\pi iy}\, u_{\mathbf{k}}\, e^{iCk_x}, \end{aligned} \tag{7.3}$$

where C is the Chern integer. The cylindrical gauge is assumed to be continuous inside the rectangle of the BZ and G_x-periodic in k_x, so it is continuous on the cylinder. This leads to the continuity of the Berry connection $\boldsymbol{\mathcal{A}}(\mathbf{k})$ on the cylinder and, hence, Gauss's theorem may be applied to the definition (7.1) to write

$$C = \frac{1}{2\pi} \oint_{\partial \mathrm{BZ}} \boldsymbol{\mathcal{A}}(\mathbf{k}) \cdot d\mathbf{k}, \tag{7.4}$$

where the boundary ∂BZ of the BZ consists of the top and bottom loops ($S^1 \oplus S^1$) of the cylinder at $k_x = -\pi$ and π. From Eq. (7.4) the consistency of the chosen gauge with the definition of the Chern number becomes obvious. That is, C in the exponent of the boundary conditions of Eqs. (7.2-7.3) is exactly the Chern number. Note that since we consider here a single isolated band, this Chern number is a gauge-invariant quantity.

7.1.2 Multiband case and individual Chern numbers

In a multiband case of an isolated group of $\mathcal{N}$ bands non-Abelian generalizations of connection (3.16) and curvature (3.17) are used [63, 64], leading to the general expression (3.18) for the Chern number.

If we now suppose that in the group of bands under consideration each of the $\mathcal{N}$ bands is isolated – that is, separated from the others by finite gaps – then the total Chern number of the subspace is just the sum

$$C = \sum_{n=1}^{\mathcal{N}} c_n \tag{7.5}$$

of the individual Chern numbers of all the bands in the subspace [122], where c_n are computed for isolated bands as described in the preceding section. Being treated in this way, each c_n is an integer. However, one might be tempted to define the quantity

$$\tilde{c}_n = \frac{1}{2\pi} \int_{BZ} d^2k \, \mathcal{F}_{nn,xy}(\mathbf{k}) \tag{7.6}$$

as the single-band contribution of band n to C. Thus defined, $\tilde{c}_n$ is *not* necessarily an integer, since it is now allowed to mix the bands by a transformation of the form of Eq. (3.5), which can change $\mathcal{F}_{nn,xy}$. Hence, in the multiband case the partial Chern contributions defined by Eq. (7.6) are not topologically invariant.

To better understand this point, consider an example of an isolated group of $\mathcal{N} = 2$ bands, where the two bands do not touch anywhere in the BZ. In this case, each of the bands has a well-defined integer Chern index $\tilde{c}$. Thus, both states can always be written in the cylindrical gauge of Eq. (7.2). Now take a simple superposition of these two states $\psi_{\mathbf{k}} = (\psi_{1\mathbf{k}} + \psi_{2\mathbf{k}})/\sqrt{2}$. One can see that once $\tilde{c}_1 \neq \tilde{c}_2$, such a state is ill-defined on a torus, in the sense $\langle \psi_{\mathbf{k}+2\pi\hat{\mathbf{y}}} | \psi_{\mathbf{k}} \rangle \neq 1$. This means that the Chern theorem does not apply to such a state, and plugging it into the Eq. (7.6) generally gives a non-integer result.

This example suggests that in certain gauges the subspace under consideration may be decomposed into the direct sum of smaller subspaces for which Chern numbers are well defined. In this particular example the gauge that naturally realizes this decomposition is the Hamiltonian gauge, that is, the gauge in which the Hamiltonian is diagonal. However, one might wonder whether such a decomposition is still possible for overlapping bands.

A QSH insulator has a nontrivial topology [1], which can be seen as an obstruction for constructing smooth Bloch functions in a gauge that respects the $\mathcal{T}$ symmetry of the Hamiltonian [27, 110, 116]. In what follows we describe a generic procedure for decomposing the occupied subspace of such an insulator into a direct sum of Chern subspaces, i.e., disentangling the occupied subspace into bands with well-defined individual integer Chern numbers c_n. We also show that each of these Chern bands may be represented in the cylindrical gauge of Eq. (7.3) with C replaced with c_n.

7.2 Decomposition into Chern subspaces

In this section we develop a general procedure for disentangling a Kramers pair of occupied states of a 2D $\mathbb{Z}_2$-insulator into two Chern bands with individual Chern numbers $c_1 = -1$ and $c_2 = 1$. The decomposition method makes heavy use of the concept of parallel transport described in Sec. 4.4. The procedure is illustrated by its application to the Kane-Mele (KM) model that is reviewed in

Sec. 3.4.3.[1] We start by using parallel transport of the Bloch states to move the gauge discontinuity to the edge of the BZ. This makes the gauge continuous on the cylinder in k-space. The next step is to apply certain gauge transformations to split the occupied subspace into a direct sum of two subspaces that are mapped onto one another by $\mathcal{T}$. We then explain how to impose the cylindrical gauge on the two disentangled bands. Since by the time of this step the bands are already continuous in the interior of the cylinder, it is only the form of the discontinuity at the edge that has to be modified. Finally, we discuss the relation of our decomposition to the previously proposed "spin Chern numbers," also discussed in Sec. 3.4.2.

7.2.1 Moving the gauge discontinuity to the BZ edge

We now consider a general model of a TR-symmetric insulator in 2D. For simplicity we consider a minimal model with two occupied bands only, since it is the Kramers pairs near the Fermi level that are responsible for a topological phase. Thus, we consider the solution of the Schrödinger equation $H(\mathbf{k})|u_{n\mathbf{k}}\rangle = E_{n\mathbf{k}}|u_{n\mathbf{k}}\rangle$ under the TR-invariance condition $\theta H(\mathbf{k})\theta^{-1} = H(-\mathbf{k})$. As was discussed above, the BZ is assumed to have been reduced to a square spanning $[0, 2\pi] \times [-\pi, \pi]$.

We start by taking two occupied states $|u_1\rangle$ and $|u_2\rangle$ resulting from numerical diagonalization at $(0, 0)$. By TR invariance, these must be Kramers-degenerate at this point. Numerical diagonalization brings random phases to both states; we accept the random phase assigned to $|u_1\rangle$, but ensure that the second state is a Kramers partner to the first by setting $|u_2\rangle = \theta|u_1\rangle$. Starting from these states we move the gauge discontinuity to the edge of the BZ in several steps.

[1] In order for the BZ of the KM model to have the required square geometry, we use orthogonal coordinates $\tilde{k}_x$ and $\tilde{k}_y$, that are related to k_x and k_y of the original model by $\tilde{k}_x = k_x/2 - \sqrt{3}k_y/2$ and $\tilde{k}_y = k_x/2 + \sqrt{3}k_y/2$. The lattice constant is assumed to be of unit length. (Entries of the Table 3.1 have $x = (\tilde{k}_x + \tilde{k}_y)/2$ and $y = (\tilde{k}_y - \tilde{k}_x)/2$.) When applying the outlined procedure to KM model, $\tilde{k}_x$ and $\tilde{k}_y$ are used instead of k_x and k_y of the text.

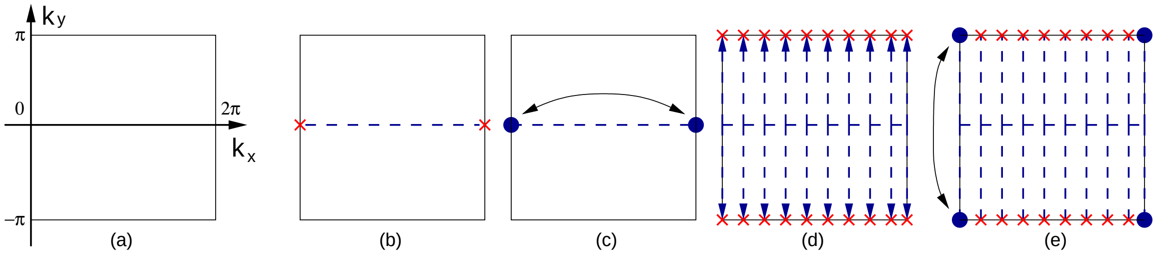

Figure 7.2: (a) BZ in k-space. (b) States are parallel transported along $k_y = 0$, but are not periodic. (c) Periodicity is restored at $k_y = 0$. (d) Parallel transport of states at all k_x from $k_y = 0$ to $k_y = \pm\pi$. (d) Periodicity is restored at $k_x = 0$, and, hence, $k_x = 2\pi$, but not at other k_x.

Parallel transport along k_x at $k_y = 0$.

As a first step of our procedure, we carry out a multiband parallel transport from $\mathbf{k} = (0,0)$ to $(2\pi,0)$ along the k_x axis. This procedure is described in detail in Sec. 4.4, but in brief it works as follows. Starting from the the two occupied states at $(0,0)$, we step along a mesh of k_x values, each time carrying out a 2×2 unitary rotation of the two states at the new k_x such that the 2×2 matrix of overlaps with the states at the previous k_x is as close as possible to the identity. The 2×2 unitary matrix U relating the states $\psi_{n\mathbf{k}}$ at $(2\pi,0)$ to those at at $(0,0)$ (i.e., the Λ matrix of Eq. (4.23)) is then constructed; its eigenvalues $\lambda_n = e^{i\phi_n}$ yield the non-Abelian Berry phases ϕ_n [63, 64]. In the present case, the $\mathcal{T}$ symmetry ensures that $\lambda_1 = \lambda_2$, so that U is just the identity times $e^{i\phi}$ where $\phi = \phi_1 = \phi_2$. Finally, the gauge discontinuity from $(2\pi,0)$ back to zero is "ironed out" by applying the gradual phase rotation $e^{-i\phi k_x/2\pi}$ to the two states at each k_x.

As a result of this procedure, we have a set of states that are smooth functions of k_x on the circular cross section of the BZ torus at $k_y = 0$, including across the seam connecting k_x=0 to $k_x = 2\pi$. This is illustrated schematically in Fig. 7.2(b-c).

Parallel transport along k_y at each k_x.

Next, at each mesh point k_x, we carry out two independent parallel-transport procedures, one from $(k_x, 0)$ to (k_x, π) along $+\hat{y}$ and another from $(k_x, 0)$ to $(k_x, -\pi)$ along $-\hat{y}$. At each new k_y point, the states are rotated by a unitary matrix so that the matrix of overlaps with the previous pair is as close to unity as possible. Starting this procedure from the line $k_y = 0$ guarantees that the states on this line remain unchanged, preserving the smoothness obtained previously. Moreover, the entire parallel-transport procedure is identical at $k_x = 0$ and $k_x = 2\pi$, ensuring that the states selected in this way are continuous across the entire seam where $k_x = 0$ has been glued to $k_x = 2\pi$. Thus, we end up with two states defined everywhere on the mesh of k points in such a way that they are smooth inside the BZ and periodic in k_x, or equivalently, smooth everywhere on the cylinder of Fig. 7.1(b). This step is illustrated in Fig. 7.2(d). The above procedure relates the states at $(k_x, -\pi)$ to those at (k_x, π) by a unitary matrix

$$V_{mn}(k_x) = \langle u_{m(k_x, k_y=-\pi)} | \, e^{2\pi i y} \, | u_{n(k_x, k_y=\pi)} \rangle, \tag{7.7}$$

which plays a role similar to Λ of Eq. (4.23). This matrix encodes the information about the gauge discontinuity that occurs on the boundary of the cylindrical BZ. Its off-diagonal elements contain information about entanglement of the two states, while the diagonal ones carry information about phase discontinuity of the states.

Restoring periodicity in k_y at $k_x = 0$.

The fact that the two states at $\mathbf{k} = 0$ form a Kramers pair guarantees that the matrix $V(k_x)$ is diagonal at $k_x = 0$ with two degenerate eigenvalues $\lambda(k_x = 0) = e^{i\varphi_0}$. (Incidentally, the same is true at $k_x = \pi$; we use this fact later.) Now we want to restore the smoothness across $k_y = \pm\pi$ at $k_x = 0$, but in such a way as to preserve the smoothness inside the cylindrical BZ. We do this by multiplying all states by a phase factor that depends smoothly on k_y only:

$$|u^{\text{new}}_{n(k_x,k_y)}\rangle = e^{-ik_y\varphi_0/2\pi}|u_{n(k_x,k_y)}\rangle \tag{7.8}$$

After this transformation, the $V(k_x)$ matrix is the identity at k_x=0. Thus, the gauge discontinuity, which has already been segregated to the edges at $k_y = \pm\pi$, has now been further excluded from the point lying at k_x=0 (or 2π) on the edge. Fig. 7.2(e) illustrates this, where red crosses on the edges represent the gauge discontinuity and the black dots indicate continuity.

Note that the entire procedure up to this point preserves the $\mathcal{T}$ symmetry, so that the states obtained so far on the BZ respect the constraints

$$\begin{aligned}\theta|u_{1\mathbf{k}}\rangle &= |u_{2-\mathbf{k}}\rangle, \\ \theta|u_{2\mathbf{k}}\rangle &= -|u_{1-\mathbf{k}}\rangle. \end{aligned} \tag{7.9}$$

This in turn implies that

$$V(-k_x) = \sigma_y\,[V(k_x)]^T\,\sigma_y \tag{7.10}$$

so that $\det[V(-k_x)] = \det[V(k_x)]$.

Removing the U(1) gauge discontinuity.

Obviously, $V(k_x) \in$ U(2), which can always be written as a U(1) phase times an SU(2) matrix. For our next step, we find it convenient to reduce $V(k_x)$ to SU(2) form by multiplying the states $|u_{n\mathbf{k}}\rangle$ by a $\mathbf{k}$-dependent phase factor. To do so, we define

$$\gamma(k_x) = \operatorname{Im}\log\det V(k_x) \tag{7.11}$$

with the branch choice that $\gamma = 0$ at k_x=0 and $\gamma(k_x)$ is a continuous function of increasing k_x. This results in $\gamma = 0$ again at $k_x = 2\pi$ because the $\mathcal{T}$ symmetry forces the total Chern number C of the two bands to be zero. Indeed, C is just given by the winding number of the U(1) → U(1) mapping from k_x to γ. This follows from

$$\begin{aligned}
2\pi C &= \int_0^{2\pi} dk_x \left[\operatorname{Tr} \mathcal{A}_{k_x}^{(k_y=-\pi)} - \operatorname{Tr} \mathcal{A}_{k_x}^{(k_y=\pi)} \right] \\
&= \int_0^{2\pi} dk_x \, \operatorname{ImTr} \left[V^\dagger \partial_{k_x} V \right] \\
&= \int_0^{2\pi} dk_x \, \partial_{k_x} \gamma(k_x) \\
&= \gamma(k_x) \Big|_0^{2\pi} \qquad (7.12)
\end{aligned}$$

after some algebra.

Thus, our next step is simply to shift the phases of all states according to

$$|u^{\text{new}}_{n(k_x,k_y)}\rangle = e^{-i\gamma(k_x)k_y/4\pi} |u_{n(k_x,k_y)}\rangle. \qquad (7.13)$$

This conserves all of the previous properties (smooth gauge inside the cylindrical BZ and on all boundaries except at $k_y = \pm\pi$). Moreover, $V(k_x = 0)$ is still the identity, but now in addition, $\det V(k_x)$ is real and positive at all k_x. That is, $V(k_x)$ has been reduced to SU(2) form. We also note that Eqs. (7.9) and (7.10) continue to hold. However, $V(k_x)$ remains off-diagonal at general k_x, thus signaling that the decomposition of the occupied subspace into the direct sum of the two TR-symmetric subspaces is not yet complete.

As noted earlier, the fact that our procedure starts from Kramers-degenerate pairs at $(k_x, k_y) = (0, 0)$ and $(\pi, 0)$ and respects $\mathcal{T}$ symmetry at all stages enforces that $V(k_x)$ must be a constant times the identity at $k_x = 0$ and $k_x = \pi$. Since

$V \in \mathrm{SU}(2)$ as well, V must be I or $-I$ at these two k_x values. Previous gauge-fixing choices insure that $V(0) = I$, but is $V(\pi) = I$ or $-I$? It can be shown that these choices correspond to the case of the Z_2 index being even or odd, respectively. Indeed, according to homotopy theory, the mapping $\mathrm{U}(1) \rightarrow \mathrm{SU}(2)$ is characterized by a Z_2 index; this is precisely the case here. In fact, the procedure up to this point can be used as an alternative to the method we presented earlier in Ref. [39] to compute the Z_2 invariant. From the numerical perspective, however, such a method does not have any significant advantages compared to the previously suggested one, apart from its straightforward geometric interpretation. In fact, for large systems it might not be very convenient to carry out all the transformations of the wavefunctions described above.

In what follows, we assume that the Z_2 index is odd.

7.2.2 Disentangling the two bands

In order to proceed, we want to make $V(k_x)$ diagonal at each k_x. When this is accomplished we will have two disentangled bands 1 and 2, although each will still have its own phase discontinuity along the boundaries at $k_y = \pm\pi$. We take a first step in this direction by taking advantage of the freedom that we had when choosing the initial representatives of the occupied subspace at $\mathbf{k} = (0, 0)$. These two states may be changed by a unitary transformation $\mathcal{U}$, which we take to belong to $\mathrm{SU}(2)$ so that the TR symmetry is fully preserved. So, we first look for the global $\mathrm{SU}(2)$ rotation that will minimize the sum of all the off-diagonal terms of the V matrices at all k_x. Once this is done, a further adjustment can be made so as to make $V(k_x)$ exactly diagonal at each k_x without losing smoothness on the cylinder. We now explain the procedure in detail.

Steepest-descent minimization of $\mathcal{V}_{\mathrm{OD}}$

Let us introduce a functional

$$\mathcal{V}_{\mathrm{OD}} = \frac{1}{N_x} \sum_{k_x} \sum_{m \neq n} |V_{mn}(k_x)|^2 \tag{7.14}$$

that is a measure of the degree to which $V(k_x)$ fails to be diagonal along the discontinuity at $k_y = \pm\pi$. The sum on k_x runs over a uniform grid of N_x mesh points. We want to use the freedom of choosing the initial pair of states at $\mathbf{k} = (0, 0)$ to minimize this functional by rotating the states at all k-points by the same unitary matrix $\mathcal{U}_0$. To do so, we consider the gradient of $\mathcal{V}_{\mathrm{OD}}$ with respect to an infinitesimal k-independent unitary transformation

$$U_{mn} = \delta_{mn} + dW_{mn}, \tag{7.15}$$

where $dW = -dW^\dagger$ for U to be unitary. A transformation of this form rotates the states according to

$$|\tilde{u}_{n\mathbf{k}}\rangle = |u_{n\mathbf{k}}\rangle + \sum_m dW_{mn} |u_{m\mathbf{k}}\rangle. \tag{7.16}$$

To first order in dW the change in $V(k_x)$ is

$$dV_{mn} = [V, dW]_{mn} \ . \tag{7.17}$$

To compute the gradient

$$G_{mn} = \left(\frac{d\mathcal{V}_{\mathrm{OD}}}{dW} \right)_{mn} = \frac{d\mathcal{V}_{\mathrm{OD}}}{dW_{nm}} \tag{7.18}$$

we note that Eq. (7.14) can be rewritten in the form

$$\mathcal{V}_{\rm OD} = \mathcal{N} - \frac{1}{N_x} \sum_{k_x} \sum_{n}^{\mathcal{N}} |V_{nn}(k_x)|^2. \tag{7.19}$$

Then, using Eq. (7.17), one can write

$$\begin{aligned} d\mathcal{V}_{\rm OD} &= -\frac{2}{N_x} \mathrm{Re} \sum_{k_x} \sum_{nm} V^*_{nn} (V_{nm} dW_{mn} - dW_{nm} V_{mn}) \\ &= -\frac{2}{N_x} \sum_{k_x} \mathrm{Re}\,\mathrm{Tr}\,[R(k_x)\, dW] \end{aligned} \tag{7.20}$$

(the k_x dependence of V is suppressed for brevity) and

$$R_{mn}(k_x) = V_{nm}[V^*_{nn} - V^*_{mm}]. \tag{7.21}$$

The second line of Eq. (7.20) is obtained by interchanging the dummy nm indices in the second term of the first line. It then follows that

$$G = \frac{1}{N_x} \sum_{k_x} \left[R(k_x) - R^\dagger(k_x) \right]. \tag{7.22}$$

We emphasize that the gradient G is independent of k_x since it generates a global unitary rotation to be applied simultaneously to all states. Also, G is not only antihermitian but also traceless, so that it generates a SU(2) unitary rotation. We now follow an iterative steepest-descent procedure, choosing a small positive damping constant β and letting $dW = -\beta G^\dagger$ (i.e, $dW = \beta G$) so that $d\mathcal{V}_{\rm OD} = \mathrm{Tr}[G\, dW] = -\beta ||G||^2$ to first order in β. We use this to update the states according to

$$|u_n^{(j+1)}\rangle = \sum_m \left[e^{\Delta W^{(j+1)}} \right]_{mn} |u_n^{(j)}\rangle \tag{7.23}$$

and the V matrices according to

$$V^{(j+1)} = \left[e^{\Delta W^{(j+1)}}\right]^{\dagger} V^{(j)} e^{\Delta W^{(j+1)}} \tag{7.24}$$

where the upper index refers to the iteration step. The iteration stops when $\mathcal{V}_{\mathrm{OD}}^{(j)} - \mathcal{V}_{\mathrm{OD}}^{(j+1)}$ stays consistently below some pre-chosen tolerance ε.

To give a flavor of how steepest descent works we give the values obtained for the KM model in the QSH regime ($\lambda_v/t = 1$, $\lambda_{SO}/t = 0.6$, $\lambda_R/t = 0.5$) with a 120×120 k-mesh, $\varepsilon = 10^{-6}$ and $\beta = 0.25$. Initially $\mathcal{V}_{\mathrm{OD}} = 0.0226$, while after minimization $\mathcal{V}_{\mathrm{OD}} = 0.0021$, so it becomes approximately ten times smaller. The crucial thing is that this final value of $\mathcal{V}_{\mathrm{OD}}$ suggests that the average off-diagonal element of V has is of order $\times 10^{-2}$, meaning that the V matrix is almost diagonal.

Note that at this stage the two subspaces are still not completely disentangled into two well-defined Chern subspaces. However, the gauge is very close to what we need. For example, the winding of $V(k_x)$ already has the necessary features: if one plots V_{11} in the complex plane as a function of k_x, one will see that it winds once around the origin in the counterclockwise direction as k_x goes from 0 to 2π, as illustrated in Fig. 7.3. Since V are not diagonal yet the trace is not the unit circle, although it is close. V_{22} winds in the opposite direction.

Diagonalization and final decomposition

Now we are in a position to make the final step in decomposition procedure. As a result of the steps above, the off-diagonal elements of the $V(k_x)$ matrices should be small compared to the diagonal ones, so that the matrices are almost diagonal. This means that $V(k_x)$ can be diagonalized by a unitary transformation $\mathcal{U}(k_x)$ that is only slightly different from the unit matrix. Since diagonalization of $V(k_x)$ does not fix the phases of the eigenvectors, and we need the phases to vary smoothly, we need an extra step to fix these phases. We do this by enforcing

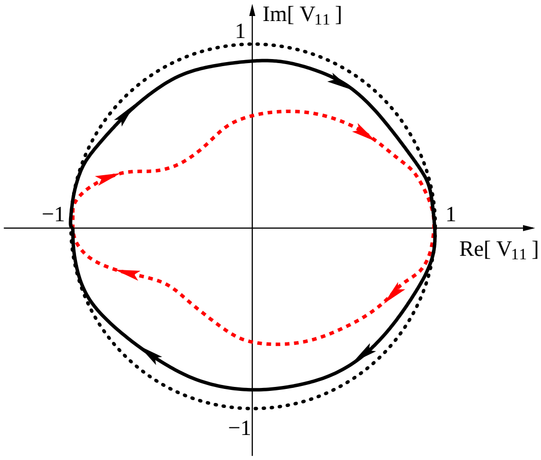

Figure 7.3: Trajectory of V_{11} in the complex plane as k_x runs across the BZ, before (red dashed line) and after (solid black line) the global $\mathcal{U}_0$ rotation that minimizes $\mathcal{V}_{\mathrm{OD}}$ for a $\mathbb{Z}_2$-odd insulator. In neither case is the graph exactly a unit circle (dotted line), but $V_{11}(0) = 1$ and $V_{11}(\pi) = -1$.

that the dominant component of each eigenvector of $V(k_x)$ is real and positive.[2] We then apply $\mathcal{U}(k_x)$ to rotate the states at all k_y for each given k_x (except at $k_x = 0$ or π, where V was already diagonal).

As a result of this step the occupied subspace has been disentangled into a direct sum of two subspaces corresponding to states $n = 1$ and 2. Moreover, they should form Kramers pairs and satisfy the constraint (7.9). Each subspace has a gauge that is smooth on the cylinder but not on the torus, since there is still a phase mismatch, corresponding to $V_{nn}(k_x)$, across the boundary at $k_y = \pm\pi$. For the Z_2-odd case this phase discontinuity can never be completely removed, since the subspaces have Chern numbers of ±1.

To check the procedure, we apply it to the KM model and compute the individual Chern numbers of the two disentangled bands. The computation is done for each band separately using the Abelian definition of Berry curvature,

[2] As an alternative, one could carry out a single-band parallel transport of the two resultant states along k_x to smooth out the random phase variations at different k_x introduced by the diagonalization procedure.

Eq. (7.1). The result is $C_1 = -1$ and $C_2 = +1$. The fact that the two states have well-defined Chern numbers is a signature of disentanglement, so that the individual Chern numbers of Eq. (7.6) have integer values ($c_1 = -c_2 = -1$). The $\mathcal{T}$ constraint of Eq. (7.9) is indeed respected at each k-point. Thus we conclude that we have succeeded in finding a decomposition of the occupied subspace into a direct sum of two Chern subspaces that are mapped onto each other by the $\mathcal{T}$ symmetry. Once again, we see that the TR-symmetric gauge for topological insulators (TIs) is discontinuous on the BZ torus.

7.2.3 Establishing a cylindrical gauge

In Sec. 7.1.1 we introduced a special "cylindrical gauge" for which the states satisfy Eq. (7.3). The defining characteristic of this special gauge is that the phase discontinuity at the cylinder boundary evolves at a *constant rate* as a function of k_x. As we shall see in Sec. 7.3, it is useful to have such a "standard gauge" enforced on the states when using them in some subsequent operations. Here we show how to extend our procedure so as to conform to the requirements of the cylindrical gauge.

As was mentioned above, the diagonal elements of $V(k_x)$ wind around zero in the complex plane in opposite directions, changing by 2π when k_x goes from 0 to 2π. Since we have carried out the diagonalization of the V matrices, we know that the V_{jj} elements follow a unit circle in the complex plane of the form $e^{i\rho_j(k_x)}$. However, the speed of this rotation given by $v_j(k_x) = d\rho_j/dk_x$ (where ρ_j remains on the same branch of the logarithm) is not constant, in contrast to the requirement of the cylindrical gauge.

To change the speed of winding of $V(k_x)$ we apply the gauge transformation

$$W(k_x, k_y) = \left[V_{\text{targ}}(k_x)V^\dagger(k_x)\right]^{k_y/2\pi} \tag{7.25}$$

to the the occupied states at each (k_x, k_y). Here

$$V_{\text{targ}}(k_x) = \begin{pmatrix} e^{ik_xc_1} & 0 \\ 0 & e^{ik_xc_2} \end{pmatrix}.$$

gives the target shape of v that corresponds to the cylindrical gauge. Note that the choice of sign should be correlated with the individual Chern number of the band it is applied to. Such a gauge transformation is obviously continuous on the cylinder and does not change the topology of the individual bands. It also preserves the $\mathcal{T}$ symmetry of the states and the relation of Eq. (7.9) is still satisfied.

We note that if the above decomposition is applied to a normal insulator (say, the KM model in the normal-insulator regime), then $c_1 = c_2 = 0$ and a smooth gauge is obtained at this step.

7.2.4 Relation to spin Chern numbers

Finally, we would like to compare our approach to disentangling $\mathbb{Z}_2$ bands into Chern bands to some other approaches suggested previously. In the work of Ref. [81] the authors suggested to associate a Chern number with each possible spin projection value. This is especially convenient when $\hat{s}_z$ is conserved; then it is natural to assign individual Chern numbers to each of the bands identified by a particular value of s_z. Such Chern numbers were called “spin Chern numbers.” For example, in the case of the KM model with no Rashba coupling (i.e., $\lambda_R = 0$), $\hat{s}_z$ is conserved and the Hamiltonian becomes block-diagonal with respect to the spin projection, allowing for well-defined spin Chern numbers. When the Rashba interaction is turned on the mirror symmetry of the model is broken and $\hat{s}_z$ is no longer conserved, thus making the original concept of a spin Chern number obscure.

This issue was clarified further by Prodan [119], who showed that even with the spin-mixing Rashba term it is possible to define spin Chern numbers by diagonalizing $\hat{s}_z$ in the occupied space of $\mathbb{Z}_2$ insulator at each $\mathbf{k}$. In other words, one diagonalizes the operator $\hat{P}_{\mathbf{k}}\hat{s}_z\hat{P}_{\mathbf{k}}$, where $\hat{P}_{\mathbf{k}}$ is the projector onto the occupied states at $\mathbf{k}$. Then, if the eigenvalues turn out to be separated by a spectral gap from one another at each value of $\mathbf{k}$, one can identify these "bands" as the desired manifolds, and carry out a unitary rotation of the original bands into these states to disentangle them. The spin Chern numbers thus defined for these bands are well defined and, in fact, correspond to the individual Chern numbers of our work. However, when the spectral gap between any two eigenvalues of the projected spin operator closes, such a decomposition becomes impossible. One could still consider some other projection operators based on mirror or other symmetries, as in Ref. [118], and use these eigenvalues in a similar way to disentangle the occupied states. However, such a method always relies on some symmetry of a particular model, and is thus not universal. The method suggested in the present work, in contrast, does not depend on any symmetries of the underlying system. Thus, we conclude that individual Chern numbers proposed in the present work are robust and arise solely from the topology of the occupied subspace of the system.

Finally, it was discussed elsewhere that the spin Chern numbers do not contain any more information than the $\mathbb{Z}_2$ invariant, because their sign can be changed without closing the insulating gap [27, 119, 123]. This is the case for individual Chern numbers as well, since obviously, one can simply change the labeling of the states by a simple unitary transformation that interchanges $|u_{1\mathbf{k}}\rangle$ with $|u_{2\mathbf{k}}\rangle$. Therefore, individual Chern numbers are merely an alternative way of describing the occupied subspace of a $\mathbb{Z}_2$ insulator in terms of disentangled bands, and do not contain any more information about the topological state of the whole system than a $\mathbb{Z}_2$ invariant alone.

7.3 Rotation into a smooth gauge

We now discuss the final step in our construction of a smooth gauge for a QSH insulator starting from the two Chern bands obtained at the previous steps. The task of unwinding the topological twists of these bands requires a unitary transformation that is also topologically nontrivial in the following sense. Obviously, a transformation that is smooth on the BZ torus, being periodic in the k_y direction, cannot make a cylindrical gauge smooth. One needs instead a unitary transformation $\mathcal{G}(\mathbf{k}) \in \mathrm{U}(2)$ that has a discontinuity on the torus that exactly cancels out the discontinuities of the cylindrical-gauge states. Of course, since the total Chern number of the whole occupied space is a topological invariant [122], the transformation will preserve the condition that the total Chern number is zero. In particular, the rotation we are looking for makes $c_1 = c_2 = 0$.

A unitary transformation that solves the problem of unwinding the two QSH bands with Chern numbers ±1 is given naturally by the solution of the Haldane model [19] of a Chern insulator (CI), or for that matter, of any two-band model of a CI. Indeed, the unitary transformation $\mathcal{G}(\mathbf{k})$ that diagonalizes the Hamiltonian in that case is one that rotates the two topologically trivial tight-binding basis states $(1,0)^T$ and $(0,1)^T$ into the eigenstates of the model. Obviously, $\mathcal{G}^{-1}(\mathbf{k}) = \mathcal{G}^{\dagger}(\mathbf{k})$ rotates the topologically nontrivial states back into the trivial ones, and thus can be used to unwind our QSH states. In order for this procedure to produce a smooth gauge, the Hamiltonian eigenstates of the CI model also have to be smoothly defined on the cylinder and obey the same cylindrical gauge of Eqs. (7.2-7.3). Assuming this has been done, the application of the resulting $\mathcal{G}^{\dagger}(\mathbf{k})$ to the QSH states defined by our procedure will finally result in a gauge that is smooth everywhere on the torus and that generates new bands with $c_1 = c_2 = 0$, as desired.

The numerical implementation of this procedure is done most conveniently by solving the CI model on the same 2D $\mathbf{k}$-space mesh as was used to solve for the QSH states. If the latter have been computed in the context of first-principles calculations or of some complex tight-binding model, then some known CI model such as the Haldane model can be used to provide the needed $\mathcal{G}(\mathbf{k})$. However, when working with a minimal 4×4 tight-binding model for a QSH system, it may be more convenient to use a 2×2 spin-up (or spin-down) block of the original 4×4 QSH model itself. After all, this already lives on the needed $\mathbf{k}$-mesh and generates bands with Chern numbers of ±1. For example, for an application to the KM model in the QSH regime ($\lambda_v/t = 1$, $\lambda_{SO}/t = 0.6$, $\lambda_R/t = 0.5$), we used the spin-up block of the original Hamiltonian and obtained two states $|u'_{i\mathbf{k}}\rangle$ with Chern numbers $c'_1 = -1$ and $c'_2 = 1$, where the hat is used to distinguish the CI quantities from the QSH ones.

As mentioned earlier, it is also necessary to bring the CI bands $|u'_{i\mathbf{k}}\rangle$ into the cylindrical gauge in order to ensure that the resulting $\mathcal{G}^\dagger(\mathbf{k})$ exactly cancels the discontinuity of the QSH bands at the edge of the cylinder. For this purpose, a parallel-transport procedure is carried out across the BZ in close analogy to what was described in Sec. 7.2.1, but now it is done in a single-band U(1) context applied to each of the CI states in turn. It is useful to refer again to Fig. 7.2. First, a parallel transport of $|u'_{1\mathbf{k}}\rangle$ is carried out along the k_x axis (with an arbitrary choice of phase at $\mathbf{k} = 0$), and a graded phase twist is applied to match phases at $k_x = 0$ and 2π as in Figs. 7.2(b-c). Then parallel transport is performed along the vertical directions as in Fig. 7.2(d), and a (k_x-independent) phase change that is graded along k_y is applied to restore continuity at the corner points of Fig. 7.2(e). This defines a phase discontinuity $V'_{11}(k_x) = \langle u'_1(k_x, -\pi)|u'_1(k_x, \pi)\rangle$ whose phase-winding rate $d\ln(V_{11})/dk_x$ is initially nonuniform, but is made uniform by the same trick as for the QSH states. The procedure is repeated for the second CI band.

The above procedure results in Chern bands obeying the cylindrical gauge as required. We can now simply form the desired unitary matrix $\mathcal{G}(\mathbf{k})$ as the 2×2 matrix whose first and second columns are filled with the column vectors $|u_1'(\mathbf{k})\rangle$ and $|u_2'(\mathbf{k})\rangle$ respectively. We emphasize again that this matrix is not topologically trivial; its coefficients are continuous on the cylinder, but not continuous across $k_y = \pm\pi$, just like the CI that has produced it. Applying $\mathcal{G}^\dagger(\mathbf{k})$ to the QSH bands constructed in Sec.7.2,

$$|\tilde{u}_{n\mathbf{k}}\rangle = \sum_m \mathcal{G}^\dagger_{mn}(\mathbf{k})|u_{m\mathbf{k}}\rangle\,, \tag{7.26}$$

we finally end up with two bands that have $c_1 = c_2 = 0$ and that span the Hilbert space defined by the original occupied bands of the QSH model. Thus, we have constructed a smooth and periodic gauge for the target $\mathbb{Z}_2$ insulator.

It should be stressed that rotation into a smooth gauge as described above breaks $\mathcal{T}$ symmetry, since $\mathcal{G}(\mathbf{k})$ results from a $\mathcal{T}$-broken CI model. Thus, the two smooth subspaces are not mapped onto each other by the $\mathcal{T}$ operator, so that $\langle \tilde{u}_{1,\mathbf{k}}|\theta|\tilde{u}_{2,-\mathbf{k}}\rangle \neq 0$ except at TR-invariant momenta $\mathbf{k} = -\mathbf{k} + \mathbf{G}$. Similarly, if Wannier functions are constructed from the Bloch spaces defined in this way, they will not form Kramers pairs [40]. Finally, we note that although the gauge is now smooth and periodic, it can be smoothed further by using this gauge as a starting point for a Wannier-function maximal-localization procedure [32].

In summary, we have demonstrated a general method for constructing a smooth gauge for a $\mathbb{Z}_2$ TI. At this final stage we start with a gauge that still respects TR symmetry, but then we carry out a unitary mixing operation that violates this symmetry in order to avoid the topological obstruction. Application to the KM model allows us to compute the $\mathbb{Z}_2$ invariant with the smooth-gauge formula of Fu and Kane [27] as discussed in the next section.

7.4 Time-reversal constraint and smooth gauge

In Ref. [27] Fu and Kane developed a theory of a $\mathbb{Z}_2$ periodic spin pump of a 1D insulating system (see Sec. 5.1.1 for details). That work established a formula (3.37) for computing the $\mathbb{Z}_2$ invariant given a smooth gauge. Here we review this result and discuss it from the perspective of the smooth gauge constructed above.

The work of Ref. [27] focuses on the periodic pumping process in 1D gapped Hamiltonians subject to the condition (5.2). Such a pump becomes $\mathcal{T}$-invariant at $t = 0$ and $t = T/2$. Assuming at the $\mathcal{T}$-invariant values of t a gauge of the form

$$\begin{aligned} \theta|u_{1k}\rangle &= e^{i\chi_k}|u_{2-k}\rangle \\ \theta|u_{2k}\rangle &= -e^{i\chi_{-k}}|u_{1-k}\rangle, \end{aligned} \tag{7.27}$$

that is smooth in k, it was shown that one can compute the $\mathbb{Z}_2$ invariant associated with the pumping process from a knowledge of the occupied states at the $\mathcal{T}$-invariant points of the pumping cycle only. However, for this purpose the gauge must be smooth on the whole torus formed by k and t [27].

Let us now look at how all this is reformulated in terms of the gauges introduced in the present chapter for a 2D system. The Hamiltonian gauge of an ordinary $\mathcal{T}$-symmetric insulating system corresponds to $\chi_k = 0$ in Eq. (7.27), and it is possible to define Bloch states in a smooth fashion on the whole torus subject to this condition. However, for a $\mathbb{Z}_2$ insulator such a constraint introduces a topological obstruction for a smooth gauge [27]. This can be understood in terms of the cylindrical gauge introduced in Sec. 7.1. Taking into account that the $\mathcal{T}$-symmetric values of the pumping parameter now correspond to $k_y = 0$ and $k_y = \pm\pi$, note that in the cylindrical gauge the $\mathcal{T}$ operator maps the states at $(k_x, k_y = 0)$ to $(-k_x, k_y = 0)$ and the states at $(k_x, k_y = \pm\pi)$ to $(-k_x, k_y = \mp\pi)$ according to Eq. (7.9). If we now take into account the boundary conditions of

Eq. (7.3) for the cylindrical gauge and use them to relate the states at (k_x, k_y) to those at $(-k_x, k_y)$, one then arrives at a relation of the form of Eq. (7.27) with

$$\chi_k = 0$$

at $k_y = 0$ and

$$\chi_k = \pm k_x C$$

at $k_y = \pm\pi$. For an ordinary insulator $C = 0$, and this obviously reduces to the standard case of $\chi_k = 0$ both at $k_y = 0$ and $k_y = \pm\pi$.

To derive an expression for the $\mathbb{Z}_2$ invariant partial polarization of Eq. (5.7) is written [27] using the gauge of Eq. (7.27) via

$$P_t^{(S)} = \frac{1}{2\pi}\left[i\int_0^\pi \langle u_{S,t,k}|\partial_k|u_{S,t,k}\rangle dk + (\chi_{t,k=\pi} - \chi_{t,k=0})\right] \tag{7.28}$$

where the index $S = 1, 2$ differentiates between the two states of a Kramers pair. This expression is U(2) invariant modulo a lattice vector ($a = 1$), provided that the transformation is globally smooth in 1D. The $\mathbb{Z}_2$ invariant was defined as [27] (see Chapter 5 for the details)

$$\nu = (P_{t=0}^{(1)} - P_{t=0}^{(2)}) - (P_{t=T/2}^{(1)} - P_{t=T/2}^{(2)}), \tag{7.29}$$

when the gauge is also smooth in t from 0 to $T/2$. With the χ_k suggested by the cylindrical gauge, and taking into account that C has opposite sign for $S = 1$ and $S = 2$, one has $P_{k_y=0}^{(1)} - P_{k_y=0}^{(2)} = 0$ and $P_{k_y=\pm\pi}^{(1)} - P_{k_y=\pm\pi}^{(2)} = \pm C$, obviously giving the correct value of the topological invariant.

As shown above, the construction of a smooth gauge starting with the cylindrical gauge proceeds by means of a unitary rotation that unwinds the gauge discontinuity of the cylindrical gauge. The unitary matrix that realizes this transformation is smooth and periodic in k_x. Thus, when establishing a smooth gauge at the $\mathcal{T}$-invariant values of k_y the gauge condition (7.27) on the 1D system is changed smoothly and, as discussed in Sec. 7.3, the smooth occupied subspaces are no longer mapped onto each other by θ. However, the $\mathcal{T}$ polarization, $P^{(1)} - P^{(2)}$, does not change under such a transformation, and as was nicely shown in Ref. [27], one can compute the $\mathbb{Z}_2$ index using the formula 3.37.

The application of the smooth-gauge construction developed in the preceding sections to the KM model in the QSH regime indeed results in the odd value for ν. The $\mathcal{T}$ constraint fixes $\text{Pf}[w]$ in (3.37) to the form $w_{12}(\mathbf{k}^*) = \pm 1$ at the $\mathcal{T}$-invariant momenta, with $|w_{12}(\mathbf{k})| < 1$ at other values of $\mathbf{k}$. In particular, our parameter choice ($\lambda_v/t = 1$, $\lambda_{SO}/t = 0.6$, $\lambda_R/t = 0.5$) results in $w_{12}(0,0) = 1$ but $w_{12}(0,\pi) = w_{12}(\pi,0) = w_{12}(\pi,\pi) = -1$, thus signaling a band inversion at $\Gamma = (0,0)$.

Chapter 8
Summary and conclusions

In this thesis we considered several aspects of band theory in the presence of topologically non-trivial bands. Here we summarize our results and discuss their possible extensions.

In Chapter 5 we have proposed a new approach for calculating topological invariants in $\mathcal{T}$-invariant systems. The method is based on following the evolution of hybrid Wannier charge centers, and is very general, being easily applicable in both tight-binding and DFT contexts. The needed ingredients are the same as those needed for the calculation of the electric polarization or the construction of maximally-localized Wannier functions, and are thus readily available in standard code packages. The presented algorithm is relatively inexpensive, however, because the analysis is confined to a small number of 2D slices of the 3D Brillouin zone. The method is easily automated and remains robust even when many bands are present. We hope that this method can help to make the search for topological phases in noncentrosymmetric materials a routine task, and that it will lead to further progress in this rapidly developing field.

In Chapter 6 we have considered the question of how to construct a Wannier representation for $\mathbb{Z}_2$ topological insulators in 2D. We have shown that the usual method based on projection onto trial functions fails because of a topological obstruction if one imposes the condition that the trial functions should come in time-reversal pairs. On the other hand, the projection method can be made to work if this condition is not imposed, resulting in WFs that do not transform into one another under time reversal.

Such a Wannier representation may have some formal disadvantages. For example, if one writes the Hamiltonian as a matrix in this Wannier representation, its time-reversal invariance is no longer transparent, and the presence of other symmetries may become less obvious as well. On the other hand, it does satisfy all the usual properties of a Wannier representation, as for example the ability to express the electric polarization in terms of the locations of the Wannier centers, and there is every reason to expect that the maximally localized WFs are still exponentially localized [36].

In Chapter 7 we have developed a general method for decomposing the occupied space of a $\mathbb{Z}_2$ insulator into a direct sum of two $\mathcal{T}$-symmetric Chern subspaces with nontrivial individual Chern numbers. We then described a general procedure for breaking the $\mathcal{T}$ symmetry between the two bands and rotating them into subspaces that are smooth everywhere on the torus. Our methods are general in the sense that they do not make use of any special symmetries or assumptions about gaps in the spectrum of spin operators. This establishes the construction of a smooth gauge for 2D topological insulators.

Bibliography

[1] C. L. Kane and E. J. Mele, "Z_2 topological order and the quantum spin Hall effect", *Phys. Rev. Lett.*, vol. 95, pp. 146802, 2005.

[2] E. Kaxiras, *Atomic and Electornic Structure of Solids*, Cambridge University Press, 2003.

[3] C. Kittel, *Introduction to Solid State Physics*, Wiley & Sons, 8th edition, 2005.

[4] J. Kohanoff, *Electornic Structure Calculations for Solids and Molecules*, Cambridge University Press, 2006.

[5] R. Martin, *Electornic Structure*, Cambridge University Press, 2004.

[6] M. V. Berry, "Quantal phase factors accompanying adiabatic changes", *Proc. R. Soc. Lon. A*, vol. 392, no. 1802, pp. 45–57, 1984.

[7] J. Zak, "Berry's phase for energy bands in solids", *Phys. Rev. Lett.*, vol. 62, no. 23, pp. 2747–2750, 1989.

[8] Di Xiao, Ming-Che Chang, and Qian Niu, "Berry phase effects on electronic properties", *Rev. Mod. Phys.*, vol. 82, pp. 1959–2007, 2010.

[9] Raffaele Resta, "Manifestations of Berry's phase in molecules and condensed matter", *J. Phys. C*, vol. 12, no. 9, pp. R107, 2000.

[10] R. D. King-Smith and David Vanderbilt, "Theory of polarization of crystalline solids", *Phys. Rev. B*, vol. 47, no. 3, pp. 1651–1654, 1993.

[11] Raffaele Resta, "Macroscopic polarization in crystalline dielectrics: the geometric phase approach", *Rev. Mod. Phys.*, vol. 66, no. 3, pp. 899–915, 1994.

[12] 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, no. 6, pp. 405–408, 1982.

[13] F. D. M. Haldane, "Berry curvature on the Fermi surface: Anomalous Hall effect as a topological Fermi-liquid property", *Phys. Rev. Lett.*, vol. 93, pp. 206602, 2004.

[14] Naoto Nagaosa, Jairo Sinova, Shigeki Onoda, A. H. MacDonald, and N. P. Ong, "Anomalous Hall effect", *Rev. Mod. Phys.*, vol. 82, pp. 1539–1592, 2010.

[15] M. Z. Hasan and C. L. Kane, "Colloquium: Topological insulators", *Rev. Mod. Phys.*, vol. 82, no. 4, pp. 3045–3067, 2010.

[16] Xiao-Liang Qi and Shou-Cheng Zhang, "Topological insulators and superconductors", *Rev. Mod. Phys.*, vol. 83, pp. 1057–1110, 2011.

[17] M. Zahid Hasan and Joel E. Moore, "Three-dimensional topological insulators", *Annual Review of Condensed Matter Physics*, vol. 2, no. 1, pp. 55–78, 2011.

[18] M Nakahara, *Geometry, Topology and Physics*, Taylor & Francis, 2003.

[19] 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, no. 18, pp. 2015–2018, 1988.

[20] Andreas P. Schnyder, Shinsei Ryu, Akira Furusaki, and Andreas W. W. Ludwig, "Classification of topological insulators and superconductors in three spatial dimensions", *Phys. Rev. B*, vol. 78, no. 19, pp. 195125, 2008.

[21] Alexei Kitaev, "Periodic table for topological insulators and superconductors", *AIP Conf. Proc.*, vol. 1134, no. 1, pp. 22–30, 2009.

[22] Andreas P. Schnyder, Shinsei Ryu, Akira Furusaki, and Andreas W. W. Ludwig, "Classification of topological insulators and superconductors", *AIP Conf. Proc.*, vol. 1134, no. 1, pp. 10–21, 2009.

[23] 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, 1982.

[24] C. L. Kane and E. J. Mele, "Quantum spin Hall effect in graphene", *Phys. Rev. Lett.*, vol. 95, pp. 226801, 2005.

[25] B. Andrei Bernevig, Taylor L. Hughes, and Shou-Cheng Zhang, "Quantum spin Hall effect and topological phase transition in HgTe quantum wells", *Science*, vol. 314, no. 5806, pp. 1757–1761, 2006.

[26] Markus Konig, Steffen Wiedmann, Christoph Brune, Andreas Roth, Hartmut Buhmann, Laurens W. Molenkamp, Xiao-Liang Qi, and Shou-Cheng Zhang, "Quantum spin Hall insulator state in HgTe quantum wells", *Science*, vol. 318, no. 5851, pp. 766–770, 2007.

[27] L. Fu and C. L. Kane, "Time reversal polarization and a Z_2 adiabatic spin pump", *Phys. Rev. B*, vol. 74, pp. 195312, 2006.

[28] L. Fu, C. L. Kane, and E. J. Mele, "Topological insulators in three dimensions", *Phys. Rev. Lett.*, vol. 98, pp. 106803, 2007.

[29] Rahul Roy, "Topological phases and the quantum spin Hall effect in three dimensions", *Phys. Rev. B*, vol. 79, no. 19, pp. 195322, 2009.

[30] J. E. Moore and L. Balents, "Topological invariants of time-reversal-invariant band structures", *Phys. Rev. B*, vol. 75, no. 12, pp. 121306, 2007.

[31] L. Fu and C. L. Kane, "Topological insulators with inversion symmetry", *Phys. Rev. B*, vol. 76, pp. 045302, 2007.

[32] Nicola Marzari and David Vanderbilt, "Maximally localized generalized Wannier functions for composite energy bands", *Phys. Rev. B*, vol. 56, no. 20, pp. 12847–12865, 1997.

[33] David Vanderbilt and R. D. King-Smith, "Electric polarization as a bulk quantity and its relation to surface charge", *Phys. Rev. B*, vol. 48, no. 7, pp. 4442–4455, 1993.

[34] Nicola Marzari, Arash A. Mostofi, Jonathan R. Yates, Ivo Souza, and David Vanderbilt, "Maximally localized Wannier functions: Theory and applications", *arXiv:1112.5411*, 2012.

[35] Arrigo Calzolari, Nicola Marzari, Ivo Souza, and Marco Buongiorno Nardelli, "*Ab initio* transport properties of nanostructures from maximally localized Wannier functions", *Phys. Rev. B*, vol. 69, pp. 035108, 2004.

[36] Christian Brouder, Gianluca Panati, Matteo Calandra, Christophe Mourougane, and Nicola Marzari, "Exponential localization of Wannier functions in insulators", *Phys. Rev. Lett.*, vol. 98, no. 4, pp. 046402, 2007.

[37] D J Thouless, "Wannier functions for magnetic sub-bands", *J. Phys. C*, vol. 17, no. 12, pp. L325, 1984.

[38] T. Thonhauser and David Vanderbilt, "Insulator/chern-insulator transition in the haldane model", *Phys. Rev. B*, vol. 74, no. 23, pp. 235111, 2006.

[39] Alexey A. Soluyanov and David Vanderbilt, "Computing topological invariants without inversion symmetry", *Phys. Rev. B*, vol. 83, pp. 235401, 2011.

[40] A. A. Soluyanov and D. Vanderbilt, "Wannier representation of Z_2 topological insulators", *Phys. Rev. B*, vol. 83, no. 3, pp. 035108, 2011.

[41] Alexey A. Soluyanov and David Vanderbilt, "Smooth gauge for topological insulators", *Phys. Rev. B*, vol. 85, pp. 115415, 2012.

[42] M. Born and J. R. Oppenheimer, "Quantum theory of molecules", *Ann. d. Physik*, vol. 84, pp. 457–484, 1927.

[43] R. Winkler, *Spin-Orbit Coupling Effects in Two-Dimensional Electron and Hole Systems*, Springer, 2003.

[44] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas", *Phys. Rev.*, vol. 136, pp. B864–B871, Nov 1964.

[45] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects", *Phys. Rev.*, vol. 140, no. 4A, pp. A1133–A1138, 1965.

[46] J. C. Slater, "A simplification of the Hartree-Fock method", *Phys. Rev.*, vol. 81, pp. 385–390, 1951.

[47] P. A. M. Dirac, "Note on the exchange phenomena in the Thomas atom", *Proc. Cam. Phil. Soc.*, vol. 26, pp. 376–385, 1930.

[48] Murray Gell-Mann and Keith A. Brueckner, "Correlation energy of an electron gas at high density", *Phys. Rev.*, vol. 106, pp. 364–368, 1957.

[49] D. M. Ceperley and B. J. Alder, “Ground state of the electron gas by a stochastic method”, *Phys. Rev. Lett.*, vol. 45, pp. 566–569, 1980.

[50] S. Goedecker, M. Teter, and J. Hutter, “Separable dual-space gaussian pseudopotentials”, *Phys. Rev. B*, vol. 54, no. 3, pp. 1703–1710, 1996.

[51] James C. Phillips and Leonard Kleinman, “New method for calculating wave functions in crystals and molecules”, *Phys. Rev.*, vol. 116, pp. 287–294, 1959.

[52] D. R. Hamann, M. Schlüter, and C. Chiang, “Norm-conserving pseudopotentials”, *Phys. Rev. Lett.*, vol. 43, pp. 1494–1497, 1979.

[53] N. Troullier and José Luriaas Martins, “Efficient pseudopotentials for plane-wave calculations”, *Phys. Rev. B*, vol. 43, pp. 1993–2006, 1991.

[54] Andrew M. Rappe, Karin M. Rabe, Efthimios Kaxiras, and J. D. Joannopoulos, “Optimized pseudopotentials”, *Phys. Rev. B*, vol. 41, pp. 1227–1230, 1990.

[55] G. B. Bachelet, D. R. Hamann, and M. Schlüter, “Pseudopotentials that work: From H to Pu”, *Phys. Rev. B*, vol. 26, pp. 4199–4228, 1982.

[56] X. Gonze et al., “Abinit : First-principles approach of materials and nanosystem properties”, *Computer Phys. Comm.*, vol. 180, pp. 2582–2615, 2009.

[57] X. Gonze et al., “A brief introduction to the abinit software package”, *Z. Kristallogr.*, vol. 220, pp. 558–562, 2005.

[58] C. Hartwigsen, S. Goedecker, and J. Hutter, “Relativistic separable dual-space gaussian pseudopotentials from H to Rn”, *Phys. Rev. B*, vol. 58, no. 7, pp. 3641–3662, 1998.

[59] Y. Aharonov and D. Bohm, "Significance of electromagnetic potentials in the quantum theory", *Phys. Rev.*, vol. 115, pp. 485–491, 1959.

[60] C. Nash and S. Sen, *Topology and Geometry for Physicists*, Dover Publications, 2011.

[61] A. Bohm, A. Mostafazadeh, H. Koizumi, Q. Niu, and J. Zwanziger, *The Geometric Phase in Quantum Systems*, Springer, 2003.

[62] A. Messiah, *Quantum Mechanics*, Dover, 1999.

[63] Frank Wilczek and A. Zee, "Appearance of gauge structure in simple dynamical systems", *Phys. Rev. Lett.*, vol. 52, pp. 2111–2114, 1984.

[64] C. Alden Mead, "The geometric phase in molecular systems", *Rev. Mod. Phys.*, vol. 64, pp. 51–85, 1992.

[65] 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, 1980.

[66] R. E. Prange and S. M. Girvin, Eds., *The Quantum Hall Effect*, Springer-Verlag, 1987.

[67] L. D. Landau and E. M. Lifshits, *Quantum Mechanics*, Butterworth Heinemann, 3rd edition, 2003.

[68] M. Kohmoto, "Topological invariant and the quantization of the Hall conductance", *Annals of Physics*, vol. 160, no. 2, pp. 343–354, 1985.

[69] Walter Kohn, "Theory of the insulating state", *Phys. Rev.*, vol. 133, no. 1A, pp. A171, 1964.

[70] Haijun Zhang, Chao-Xing Liu, Xiao-Liang Qi, Xi Dai, Zhong Fang, and Shou-Cheng Zhang, "Topological insulators in Bi_2Se_3, Bi_2Te_3 and Sb_2Te_3

with a single Dirac cone on the surface", *Nature Phys.*, vol. 5, no. 6, pp. 438–442, 2009.

[71] S. Chadov, X. L. Qi, J. Kuebler, G. H. Fecher, C. Felser, and S. C. Zhang, "Tunable multifunctional topological insulators in ternary Heusler compounds", *Nature Materials*, vol. 9, no. 7, pp. 541–545, 2010.

[72] Hai-Jun Zhang, Stanislav Chadov, Lukas Muechler, Binghai Yan, Xiao-Liang Qi, Juergen Kuebler, Shou-Cheng Zhang, and Claudia Felser, "Topological insulators in ternary compounds with a honeycomb lattice", *Phys. Rev. Lett.*, vol. 106, no. 15, 2011.

[73] Binghai Yan, Lukas Müchler, Xiao-Liang Qi, Shou-Cheng Zhang, and Claudia Felser, "Topological insulators in filled skutterudites", *Phys. Rev. B*, vol. 85, pp. 165125, 2012.

[74] Hsin 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, pp. 036404, 2010.

[75] Maxim Dzero, Kai Sun, Victor Galitski, and Piers Coleman, "Topological kondo insulators", *Phys. Rev. Lett.*, vol. 104, pp. 106408, 2010.

[76] D. Hsieh, D. Qian, L. Wray, Y. Xia, Y. S. Hor, R. J. Cava, and M. Z. Hasan, "A topological Dirac insulator in a quantum spin Hall phase", *Nature*, vol. 452, no. 7190, pp. 970–974, 2008.

[77] D. Hsieh, Y. Xia, L. Wray, D. Qian, A. Pal, J. H. Dil, J. Osterwalder, F. Meier, G. Bihlmayer, C. L. Kane, Y. S. Hor, R. J. Cava, and M. Z. Hasan, "Observation of Unconventional Quantum Spin Textures in Topological Insulators ", *Science*, vol. 323, no. 5916, pp. 919–922, 2009.

[78] Y. Xia, D. Qian, D. Hsieh, L. Wray, A. Pal, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan, "Observation of a large-gap topological-insulator class with a single Dirac cone on the surface", *Nature Phys.*, vol. 5, no. 6, pp. 398–402, 2009.

[79] 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.

[80] Takafumi Sato, Kouji Segawa, Hua Guo, Katsuaki Sugawara, Seigo Souma, Takashi Takahashi, and Yoichi Ando, "Direct evidence for the Dirac-cone topological surface states in the ternary chalcogenide $TlBiSe_2$", *Phys. Rev. Lett.*, vol. 105, pp. 136802, 2010.

[81] D. N. Sheng, Z. Y. Weng, L. Sheng, and F. D. M. Haldane, "Quantum spin-hall effect and topologically invariant chern numbers", *Phys. Rev. Lett.*, vol. 97, no. 3, pp. 036808, 2006.

[82] Igor Žutić, Jaroslav Fabian, and S. Das Sarma, "Spintronics: Fundamentals and applications", *Rev. Mod. Phys.*, vol. 76, pp. 323–410, 2004.

[83] Y. A. Bychkov and E. I. Rashba, "Properties of a 2D electron-gas with lifted spectral degeneracy", *JETP Lett.*, vol. 39, no. 2, pp. 78–81, 1984.

[84] Takahiro Fukui and Yasuhiro Hatsugai, "Quantum spin Hall effect in three dimensional materials: Lattice computation of Z_2 topological invariants and its application to Bi and Sb", *J. Phys. Soc. Jpn.*, vol. 76, no. 5, pp. 053702, 2007.

[85] Bryan Leung and Emil Prodan, "Effect of strong disorder in a three-dimensional topological insulator: Phase diagram and maps of the Z_2 invariant", *Phys. Rev. B*, vol. 85, pp. 205136, 2012.

[86] S. Ryu, C. Mudry, A. W. W. Ludwig, and A. Furusaki, "Global phase diagram of two-dimensional dirac fermions in random potentials", *Phys. Rev. B*, vol. 85, pp. 235115, 2012.

[87] P. W. Anderson, "Absence of diffusion in certain random lattices", *Phys. Rev.*, vol. 109, pp. 1492–1505, 1958.

[88] E. Abrahams, P. W. Anderson, D. C. Licciardello, and T. V. Ramakrishnan, "Scaling theory of localization: Absence of quantum diffusion in two dimensions", *Phys. Rev. Lett.*, vol. 42, pp. 673–676, 1979.

[89] Roger S. K. Mong, Jens H. Bardarson, and Joel E. Moore, "Quantum transport and two-parameter scaling at the surface of a weak topological insulator", *Phys. Rev. Lett.*, vol. 108, pp. 076804, 2012.

[90] Zohar Ringel, Yaacov E. Kraus, and Ady Stern, "Strong side of weak topological insulators", *Phys. Rev. B*, vol. 86, pp. 045102, 2012.

[91] Andrew M. Essin, Joel E. Moore, and David Vanderbilt, "Magnetoelectric polarizability and axion electrodynamics in crystalline insulators", *Phys. Rev. Lett.*, vol. 102, pp. 146805, 2009.

[92] Xiao-Liang Qi, Taylor L. Hughes, and Shou-Cheng Zhang, "Topological field theory of time-reversal invariant insulators", *Phys. Rev. B*, vol. 78, pp. 195424, 2008.

[93] Joseph Maciejko, Xiao-Liang Qi, H. Dennis Drew, and Shou-Cheng Zhang, "Topological quantization in units of the fine structure constant", *Phys. Rev. Lett.*, vol. 105, pp. 166803, 2010.

[94] Wang-Kong Tse and A. H. MacDonald, "Giant magneto-optical Kerr effect and universal Faraday effect in thin-film topological insulators", *Phys. Rev. Lett.*, vol. 105, pp. 057401, 2010.

[95] Xiao-Liang Qi, Rundong Li, Jiadong Zang, and Shou-Cheng Zhang, "Inducing a magnetic monopole with topological surface states", *Science*, vol. 323, no. 5918, pp. 1184–1187, 2009.

[96] Liang Fu and C. L. Kane, "Superconducting proximity effect and majorana fermions at the surface of a topological insulator", *Phys. Rev. Lett.*, vol. 100, pp. 096407, 2008.

[97] Raffaele Resta and Sandro Sorella, "Electron localization in the insulating state", *Phys. Rev. Lett.*, vol. 82, no. 2, pp. 370–373, 1999.

[98] Raffaele Resta, "Why are insulators insulating and metals conducting?", *J. Phys. C*, vol. 14, no. 20, pp. R625, 2002.

[99] Gregory H. Wannier, "The structure of electronic excitation levels in insulating crystals", *Phys. Rev.*, vol. 52, pp. 191–197, 1937.

[100] W. Kohn, "Analytic properties of bloch waves and Wannier functions", *Phys. Rev.*, vol. 115, no. 4, pp. 809–821, 1959.

[101] Jacques Des Cloizeaux, "Analytical properties of n-dimensional energy bands and Wannier functions", *Phys. Rev.*, vol. 135, pp. A698–A707, 1964.

[102] G. Nenciu, "Dynamics of band electrons in electric and magnetic fields: rigorous justification of the effective Hamiltonians", *Rev. Mod. Phys.*, vol. 63, pp. 91–127, 1991.

[103] S. Kivelson, "Wannier functions in one-dimensional disordered systems: Application to fractionally charged solitons", *Phys. Rev. B*, vol. 26, no. 8, pp. 4269–4277, 1982.

[104] Claudia Sgiarovello, Maria Peressi, and Raffaele Resta, "Electron localization in the insulating state: Application to crystalline semiconductors", *Phys. Rev. B*, vol. 64, pp. 115202, 2001.

[105] E. I. Blount, “Formalisms of band theory”, in *Solid State Physics*, F. Seitz and D. Turnbull, Eds., vol. 13, pp. 305–372. Academic Press, 1962.

[106] Sinisa Coh and David Vanderbilt, “Electric polarization in a Chern insulator”, *Phys. Rev. Lett.*, vol. 102, no. 10, pp. 107603, 2009.

[107] Di Xiao, Yugui Yao, Wanxiang Feng, Jun Wen, Wenguang Zhu, Xing-Qiu Chen, G. Malcolm Stocks, and Zhenyu Zhang, “Half-Heusler compounds as a new class of three-dimensional topological insulators”, *Phys. Rev. Lett.*, vol. 105, no. 9, pp. 096404, 2010.

[108] W. Feng, D. Xiao, J. Ding, and Y. Yao, “Three-dimensional topological insulators in I-III-VI_2 and II-IV-V_2 chalcopyrite semiconductors”, *Phys. Rev. Lett.*, vol. 106, no. 1, pp. 016402, 2011.

[109] M. Wada, S. Murakami, F. Freimuth, and G. Bihlmayer, “Localized edge states in two-dimensional topological insulators: Ultrathin bi films”, *Phys. Rev. B*, vol. 83, no. 12, pp. 121310, 2011.

[110] Rahul Roy, “Z_2 classification of quantum spin Hall systems: An approach using time-reversal invariance”, *Phys. Rev. B*, vol. 79, no. 19, pp. 195321, 2009.

[111] Zohar Ringel and Yaacov E. Kraus, “Determining topological order from a local ground-state correlation function”, *Phys. Rev. B*, vol. 83, pp. 245115, 2011.

[112] X. Gonze, J.-P. Michenaud, and J.-P. Vigneron, “First-principles study of As, Sb, and Bi electronic properties”, *Phys. Rev. B*, vol. 41, no. 17, pp. 11827, 1990.

[113] J. R. Wiese and L. Muldawer, “Lattice constants of Bi_2Te_3-Bi_2Se_3 solid solution alloys”, *J. Phys. Chem. Solids*, vol. 15, no. 1-2, pp. 13 – 16, 1960.

[114] Akifumi Onodera, Ichiro Sakamoto, Yasuhiko Fujii, Nobuo Mori, and Shunji Sugai, "Structural and electrical properties of GeSe and GeTe at high pressure", *Phys. Rev. B*, vol. 56, no. 13, pp. 7935–7941, 1997.

[115] Xi Dai, Taylor L. Hughes, Xiao-Liang Qi, Zhong Fang, and Shou-Cheng Zhang, "Helical edge and surface states in HgTe quantum wells and bulk insulators", *Phys. Rev. B*, vol. 77, no. 12, pp. 125319, 2008.

[116] T. A. Loring and M. B. Hastings, "Disordered topological insulators via C*-algebras", *EPL*, vol. 92, no. 6, pp. 67004, 2010.

[117] Xifan Wu, Oswaldo Diéguez, Karin M. Rabe, and David Vanderbilt, "Wannier-based definition of layer polarizations in perovskite superlattices", *Phys. Rev. Lett.*, vol. 97, no. 10, pp. 107602, 2006.

[118] Jeffrey C. Y. Teo, Liang Fu, and C. L. Kane, "Surface states and topological invariants in three-dimensional topological insulators: Application to Bi$_1$-xSb$_x$", *Phys. Rev. B*, vol. 78, pp. 045426, 2008.

[119] Emil Prodan, "Robustness of the spin-chern number", *Phys. Rev. B*, vol. 80, pp. 125327, 2009.

[120] Rui Yu, Xiao Liang Qi, Andrei Bernevig, Zhong Fang, and Xi Dai, "Equivalent expression of Z$_2$ topological invariant for band insulators using the non-abelian berry connection", *Phys. Rev. B*, vol. 84, pp. 075119, 2011.

[121] Emil Prodan, "Manifestly gauge-independent formulations of the Z$_2$ invariants", *Phys. Rev. B*, vol. 83, pp. 235115, 2011.

[122] J. E. Avron, R. Seiler, and B. Simon, "Homotopy and quantization in condensed matter physics", *Phys. Rev. Lett.*, vol. 51, pp. 51–53, 1983.

[123] Takahiro Fukui and Yasuhiro Hatsugai, "Topological aspects of the quantum spin-Hall effect in graphene: Z_2 topological order and spin chern number", *Phys. Rev. B*, vol. 75, pp. 121403, 2007.

Vita

Alexey A. Soluyanov

2012 Ph. D. in Physics, Rutgers University

2004-07 M. Sc. in Physics from St. Petersburg State University, Russia

2000-04 B. Sc. in Physics from St. Petersburg State University, Russia

2009-2012 Graduate assistant, Department of Physics and Astronomy, Rutgers University

2008-2009 Teaching assistant, Department of Physics and Astronomy, Rutgers University

2007-2008 Graduate fellow, Department of Physics and Astronomy, Rutgers University