Topological Insulators: Tight-Binding Models and Surface States
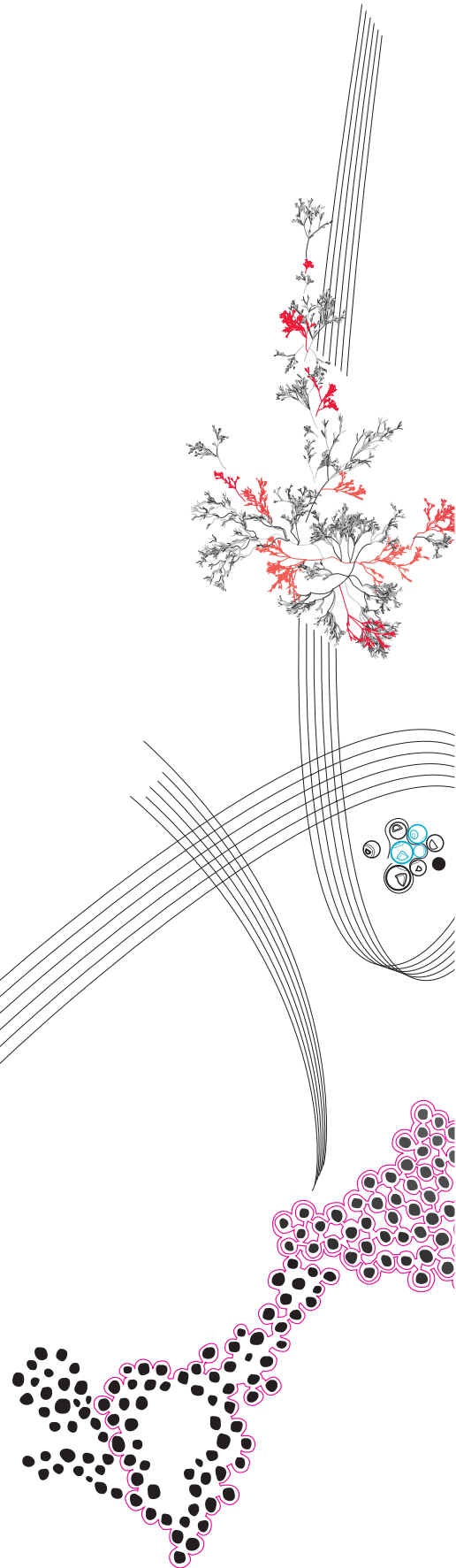
Mick Jan-Albert van Vliet

Supervisors:
prof. dr. G.H.L.A. Brocks
dr. M. Schlottbom

Second Examiners:
prof. dr. ir. A. Brinkman
prof. dr. ir. B.J. Geurts

June, 2019

UNIVERSITY OF TWENTE.

Preface

The present text is a bachelor thesis on the topic of topological insulators, written in the period April-June 2019 within the Computational Materials Science group as part of the bachelor assignment for the Bachelor Degree in Applied Physics and Applied Mathematics.

Topological Insulators:
Tight-Binding Models and Surface States

Mick J. van Vliet[*]

June, 2019

Abstract

Topological insulators are relatively recently discovered phases of quantum matter which have exotic electronic properties that have attracted an enormous amount of theoretical and experimental interest in the field of condensed matter physics. They are characterized by being electronically insulating in the bulk, while hosting topologically protected surface or edge states, where the electrons behave like massless particles. The presence of a topological insulator phase is encoded in a topological invariant taking values in $\mathbb{Z}_2$, referred to as the $\mathbb{Z}_2$-index. We compute the $\mathbb{Z}_2$-index of bismuth selenide (Bi_2Se_3) using two different methods. Subsequently, the presence of surface states in Bi_2Se_3 is investigated using the method of surface Green's functions. The results confirm that this material is a topological insulator. The relation between the $\mathbb{Z}_2$-index and the existence of surface states is described by a result called the bulk-boundary correspondence, and a proof of this result is reviewed.

Keywords: condensed matter physics, topology, topological insulators, tight-binding, surface states, bulk-boundary correspondence

[*]Email: m.j.vanvliet@student.utwente.nl

1

Contents

1 Introduction

One of the main themes of condensed matter physics is the classification of matter around us into phases of matter and the phase transitions that separate these different phases. For a long time, phases of matter have been understood in terms of certain symmetries of systems that drastically change when a phase transition occurs. Most people are familiar with the basic classification into solids, liquids and gases. Solid materials can be further classified based on their electronic properties, such as whether a given solid can conduct electricity or not. This property subdivides solids into conductors and insulators, but this is not the full story. It turns out that there is a different type of phase of matter that is not based on symmetry, but on topology.

Topology is the branch of mathematics that studies the structure of spaces. As topologists we are mainly interested in whether two given spaces are equivalent from a topological view or not, where this equivalence intuitively means that one can deform one space continuously into the other space. Such spaces are said to be homeomorphic. The main approach to study the topological structure of spaces is by assigning properties that are preserved when a space is continuously deformed. Such properties are called topological invariants, and an enormous range of topological invariants are known, in fact infinitely many. An intuitive example is the number of holes of a surface, while a more elaborate example is the number of inequivalent ways in which one can tie a loop in a space, which is encoded in a property called the fundamental group.

Because space is the place in which physics happens, it is natural to expect that topology would play some role in physical theories. And indeed, many areas of physics, mainly within high-energy physics, fundamentally rely, although somewhat implicitly, on the notion of topology. Examples are general relativity, theoretical particle physics and string theory. It is however only relatively recent that the use of topology has emerged in condensed matter physics as well. It turns out that one can meaningfully assign topological invariants to physical systems in the same way as one does for topological spaces. The topological invariant of a system defines its phase, and this type of phase of matter is called a topological phase. The invariant can only change if the system undergoes a so-called topological phase transition. The discovery of topological phases was quite revolutionary, and in fact the 2016 Nobel prize in physics has been awarded to Thouless, Haldane and Kosterlitz for their discovery of topological phases of matter [2], [4].

One of these topological phases is the topological insulator. This type of topological phase of quantum matter has been theoretically predicted and experimentally observed over the past thirty years, and the emergence of topological insulators has attracted an enormous amount of interest from both experimental and theoretical physicists [3], [4], [14]–[16]. Solid materials in a topological insulator phase are characterized by being electronically insulating in the bulk of the material, whereas they admit a flow of current over the surface. The electronic states that carry this current are topologically protected, in the sense that a topological phase transition is required to remove their existence. This means that the presence of these special surface states are robust against disorder and defects in the material structure, which makes them highly interesting for applications.

A specific type of topological insulator, namely the time-reversal symmetric topological insulator, is the main topic of this thesis. These topological insulators are characterized by a topological invariant that takes values in $\mathbb{Z}_2$, and as such this invariant is usually referred to as the $\mathbb{Z}_2$-index. Because $\mathbb{Z}_2$ has two elements, this type of topological insulator distinguishes between two phases, sometimes called the trivial phase and the topological phase. The computation of the $\mathbb{Z}_2$-index for a real material, namely bismuth selenide (Bi_2Se_3), is one aspect of this work.

Before the $\mathbb{Z}_2$-index of a given material can be computed one needs to have a description of the electronic structure of the material. Fortunately, topological insulators can be understood within the framework of single-particle quantum mechanics. We will model the electronic structure of bismuth selenide using the tight-binding method, which we will review in section 2. This section also contains a review of crystal structures, band theory, and the representation of symmetries in tight-binding models.

In section 3 the relevant theory of time-reversal symmetric topological insulators is discussed. We give a physical explanation of the meaning of the $\mathbb{Z}_2$-index from a surface perspective and from a bulk perspective, and consider two methods of its computation. The first method, due to Fu and Kane [11], is relatively simple but restricted to materials with inversion symmetry. The second method, due to Soluyanov and Vanderbilt [18], is more general but also more technical to implement. We also discuss a simple tight-binding model of a topological insulator to illustrate these concepts, called the Bernevig-Hughes-Zhang model [9], [22].

We then construct a tight-binding model for Bi_2Se_3, from which we calculate the corresponding band structure. This is done in section 4. The band structure is then used to compute the $\mathbb{Z}_2$-index of this material by applying the two methods mentioned above, showing that it is a topological insulator.

As we discussed, a non-zero $\mathbb{Z}_2$-index indicates the existence of topologically protected surface or edge states. In section 5 we show that the systems considered in the preceding sections indeed host topologically protected surface states. We do this by considering a semi-infinite lattice and calculating the associated density of states at the surface using the formalism of surface Green's functions, following [8] and [23].

The fact that a topological invariant which is computed purely from a bulk system, which in principle has no surface, has implications for phenomena that take place at the surface of a material is a non-trivial fact. The theorem that provides the link between the bulk and the boundary of a system is known as the bulk-boundary correspondence. From a mathematical point of view this is a deep result, and most proofs rely on K-theory and related tools [24]. K-theory is a theory in which topological invariants of spaces are studied in terms of vector bundles. It is from a mathematical point of view a natural tool to study topological insulators, as they are closely related to the topology of vector bundles. We will however not take this approach in this thesis. In section 6 we review a more concrete proof due to Graf and Porta [19] in the context of tight-binding models of two-dimensional systems. The full proof of this bulk-boundary correspondence is quite lengthy, so only the main steps of the proof are outlined.

Because quantum mechanics forms the foundation of our study of topological insulators, we give a brief review of quantum mechanics in appendix A. The reader who is not familiar with quantum mechanics is recommended to read this appendix before proceeding with section 2.

Before we begin, it should be remarked that topological insulators form a vast subject with many interesting aspects and different points of view (theoretical, experimental and mathematical). It is also a challenging topic which requires a considerable amount of background material. This text is intended for both physicists and mathematicians, and hence no prior knowledge of condensed matter physics is assumed. To keep the thesis moderate in size, many topics that naturally belong to a general discussion of topological insulators did not get a place in this text. Some examples of such aspects are Berry phases, the quantum Hall effect, Chern insulators and the Chern index, the quantum spin Hall effect, the role of the Dirac equation and the mathematical bulk-boundary correspondence formulated in terms of K-theory.

2 Band Theory and the Tight-Binding Method

In this section we review the theory that will form the basis of our discussion of topological insulators. We begin by briefly discussing lattices and crystal structures. We then discuss two versions of an important result for periodic quantum systems called Bloch's theorem, which exploits the translational symmetry of the Hamiltonian to reduce the full problem to a collection of simpler problems. Then we introduce an important model for the electronic structure of crystalline solids, known as the tight-binding method. As a foundation for the later sections we also pay attention to the representation of symmetries in tight-binding models. We conclude with a discussion of what it means for two quantum system to be topologically equivalent based on the notion of adiabatic continuity. For the reader who is familiar with these notions it suffices to scan this section.

2.1 Crystal Structures and Translational Invariance

Many solid materials are crystalline in nature, which means that their constituent atoms are ordered in regular and repeating patterns at the microscopic scale. These patterns are often arranged periodically, and hence form a so-called crystal structure. If one wants to describe a sample of a crystalline solid of macroscopic size, it is a powerful idealization to assume that this crystal extends infinitely far in all directions so that the system acquires a certain translational symmetry. Macroscopic samples typically consist of $10^{20} - 10^{24}$ atoms, which means that for an electron in the bulk of the sample, this is a very good approximation. On the other hand, if one is interested in effects taking place at the edge of the sample, one has to take a different approach. The electronic structure inside the bulk and at the boundary are not entirely unrelated however, as there is an important theorem for topological insulators known as the bulk-boundary correspondence. This result roughly states that the $\mathbb{Z}_2$-index determines the phenomena that take place at the surface, and it is discussed further in section 3. We begin by introducing a mathematical description of crystal lattices. For a more detailed description we refer to [5], [25].

We consider a d-dimensional crystal[1] to be a discrete subset $\mathcal{C} \subset \mathbb{R}^d$ which is invariant under the action of a group B consisting of translations by vectors lying on a certain lattice known as the Bravais lattice. Points on the crystal $\mathcal{C}$ are called sites, and typically these sites correspond to the locations of atoms. By definition, the Bravais lattice has the form

$$B = \mathbf{v}_1\mathbb{Z} \oplus \cdots \oplus \mathbf{v}_d\mathbb{Z} \subset \mathbb{R}^d,$$

where $\mathbf{v}_1, \ldots, \mathbf{v}_d$ are d linearly independent vectors in $\mathbb{R}^d$. By this notation we mean that vectors on the Bravais lattice are linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_d$ with integral coefficients, so they are of the form

$$\mathbf{R} = n_1\mathbf{v}_1 + \ldots + n_d\mathbf{v}_d,$$

for integers $n_1, \ldots, n_d \in \mathbb{Z}$. The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_d$ are called primitive vectors. As is usually done in condensed matter physics, we will interchangeably use the term Bravais lattice to refer to the group of translations as well as the underlying set of points in space. An example of a two-dimensional Bravais lattice is shown in figure 1.

---

[1]We write d for generality, but for our purposes we are interested only in $d = 1, 2, 3$.

Because of the translational invariance of the crystal, its structure is fully specified by giving the locations of sites of the crystal around one of the Bravais lattice sites. A region of space around a Bravais lattice site which tessellates all space when translated by the elements of B is known as a primitive unit cell. Hence, a full description of a crystal $\mathcal{C}$ consists of a Bravais lattice together with a specification of the sites in a primitive unit cell. This is illustrated in figure 1, which shows the crystal structure of graphene, a two-dimensional material.
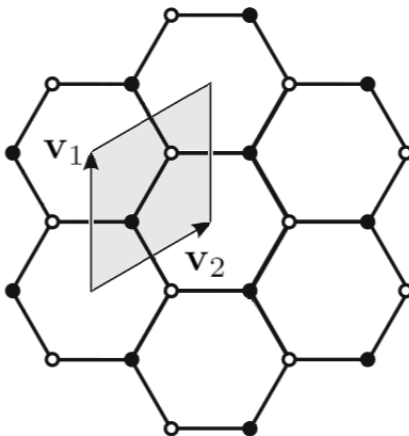
FIGURE 1: Crystal structure of graphene. The primitive vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ span a possible Bravais lattice for this crystal structure. The shaded region indicates a primitive unit cell. Black and white dots indicate the inequivalent sites of the atoms, where equivalent means that the sites are related by a Bravais lattice vector. Adapted from [19].

Instead of taking an infinite lattice, it is customary to consider a finite lattice with Born-von Karman periodic boundary conditions imposed [22]. Physically, this corresponds to taking a finite system and attaching the opposite endpoints of the lattice, so that for instance a chain of atoms becomes a ring of atoms, and a sheet of atoms takes the form of a torus. Using these boundary conditions, one does not have to deal with non-normalizable states. The physical argument for accepting these boundary conditions is that the bulk properties of a system should not depend on which boundary conditions are chosen at the edge [5]. Mathematically, one could say that the underlying Bravais lattice of such a system is a direct sum of finite cyclic groups, or

$$B = \mathbf{v}_1 \mathbb{Z}_{N_1} \oplus \cdots \oplus \mathbf{v}_d \mathbb{Z}_{N_d},$$

where N_i is the number of lattice sites in the ith direction for $i = 1, \ldots, d$. In this way, $\mathcal{N} = \prod_{i=1}^{d} N_i$ is the number of unit cells in the crystal. This description is more convenient because instead of integrals one can deal with finite sums when working with Fourier transforms. It is no longer natural to embed such a finite Bravais lattice in $\mathbb{R}^d$, and instead one embeds it in a box with periodic boundary conditions, or equivalently a d-dimensional torus, which we denote by $\mathbb{T}^d$.

For a given infinite Bravais lattice $B = \bigoplus_{i=1}^{d} \mathbf{v}_i \mathbb{Z} \subset \mathbb{R}^d$, a central concept that can be defined is the so-called reciprocal lattice or dual lattice B^*, given by all vectors $\mathbf{G} \in \mathbb{R}^d$ for which

$$\mathbf{G} \cdot \mathbf{R} \in 2\pi\mathbb{Z}, \quad \text{for any } \mathbf{R} \in B.$$

The reciprocal lattice can be constructed directly from the primitive lattice vectors $\mathbf{v}_1, \ldots, \mathbf{v}_d$ by taking the dual basis $\mathbf{w}_1, \ldots, \mathbf{w}_d$ to these vectors, which satisfy

$$\mathbf{w}_i \cdot \mathbf{v}_j = 2\pi \delta_{ij},$$

with δ_{ij} the Kronecker delta. The reciprocal lattice is then given by

$$B^* = \mathbf{w}_1 \mathbb{Z} \oplus \cdots \oplus \mathbf{w}_d \mathbb{Z}.$$

For the three-dimensional case, the dual basis vectors $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ can be obtained directly using

$$\mathbf{w}_1 = 2\pi \frac{\mathbf{v}_2 \times \mathbf{v}_3}{\mathbf{v}_1 \cdot (\mathbf{v}_2 \times \mathbf{v}_3)},$$
$$\mathbf{w}_2 = 2\pi \frac{\mathbf{v}_3 \times \mathbf{v}_1}{\mathbf{v}_1 \cdot (\mathbf{v}_2 \times \mathbf{v}_3)},$$
$$\mathbf{w}_3 = 2\pi \frac{\mathbf{v}_1 \times \mathbf{v}_2}{\mathbf{v}_1 \cdot (\mathbf{v}_2 \times \mathbf{v}_3)}.$$

The reciprocal lattice is defined similarly for a finite lattice $B = \bigoplus_{i=1}^d \mathbf{v}_i \mathbb{Z}_{N_i} \subset T^d$.

We now turn to an important result for periodic systems called Bloch's theorem. We first state the version for continuous systems, where the states are wavefunctions. For points $\mathbf{R} \in B$, we denote the translation operator by the vector $\mathbf{R}$ by $T_{\mathbf{R}}$, which acts on functions f in $L^2(\mathbb{R}^d, \mathbb{C})$ or $L^2(\mathbb{T}^d, \mathbb{C})$ by

$$T_{\mathbf{R}} f(\mathbf{r}) = f(\mathbf{r} - \mathbf{R}).$$

We remark that we treat the coordinates on the torus as periodic coordinates in $\mathbb{R}^d$.

Theorem [Bloch, continuous version]. Consider a quantum system modelled on the Hilbert space $L^2(\mathbb{T}^d, \mathbb{C})$ with a Hamiltonian H satisfying the translational invariance condition

$$H = T_{\mathbf{R}} H T_{-\mathbf{R}} \tag{1}$$

for any $\mathbf{R}$ in a Bravais lattice B. Then H and each $T_{\mathbf{R}}$ can be simultaneously diagonalized, and the common eigenfunctions can be chosen to be Bloch waves, which are functions of the form

$$\psi_{n,\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{n,\mathbf{k}}(\mathbf{r}), \tag{2}$$

where $\mathbf{k} \in \mathbb{R}^d$ is called the crystal momentum and $u_{n,\mathbf{k}}$ is a periodic function with the periodicity of the Bravais lattice.

Here n is an index labelling the eigenstates for a given $\mathbf{k}$, referred to as the band index. Typically, the condition of equation (1) arises from a periodic potential, which is the case for crystals. Another way to state the defining property of a Bloch wave is

$$T_{\mathbf{R}} \psi_{n,\mathbf{k}}(\mathbf{r}) = \psi_{n,\mathbf{k}}(\mathbf{r} - \mathbf{R}) = e^{-i\mathbf{k}\cdot\mathbf{R}} \psi_{n,\mathbf{k}}(\mathbf{r}).$$

From this condition we see that if $\mathbf{G} \in B^*$ is a reciprocal lattice vector, then

$$e^{-i(\mathbf{k}+\mathbf{G})\cdot\mathbf{R}} = e^{-i\mathbf{k}\cdot\mathbf{R}},$$

so Bloch waves with crystal momenta that differ by a reciprocal lattice vector essentially describe the same Bloch wave. Thus, we can identify the crystal momenta $\mathbf{k}$ and $\mathbf{k} + \mathbf{G}$ for all $\mathbf{G} \in B^*$. Under this identification, we thus only consider $\mathbf{k}$ in the quotient space $\mathbb{R}^d/B^*$. From basic topology we know that this quotient is a d-dimensional torus. We denote it by

$$\mathfrak{B} = \mathbb{R}^d/B^*$$

and call it the Brillouin zone[2]. For each $\mathbf{k} \in \mathfrak{B}$, we can focus on the part of the wavefunction that is periodic on the Bravais lattice, cf. equation (2). These functions are eigenfunctions of the so-called Bloch Hamiltonian, defined by

$$H(\mathbf{k}) = e^{-i\mathbf{k}\cdot\mathbf{r}} H e^{i\mathbf{k}\cdot\mathbf{r}},$$

which acts on the Hilbert space consisting of B-periodic functions on the crystal. We denote the spectra of the Bloch Hamiltonians by

$$\{\varepsilon_n(\mathbf{k})\}_{n\in J} = \sigma(H(\mathbf{k})),$$

where J is some indexing set. The energy eigenvalues $\{\varepsilon_n(\mathbf{k})\}$ are the essence of the description of the electronic structure of solids. If one plots the eigenvalue branches $\{\varepsilon_n(\mathbf{k})\}$, which depend continuously on $\mathbf{k}$, as a function of $\mathbf{k}$ in the Brillouin zone $\mathfrak{B}$ for a given material, one obtains a graph called the *band structure* of that material. In order to visualize the band structure of higher-dimensional crystals, it is customary to plot the energies $\{\varepsilon_n(\mathbf{k})\}$ for $\mathbf{k}$ along a prescribed path in the Brillouin zone, where the choice of path depends on the crystal structure of the material. An example of a band structure is shown in figure 2.
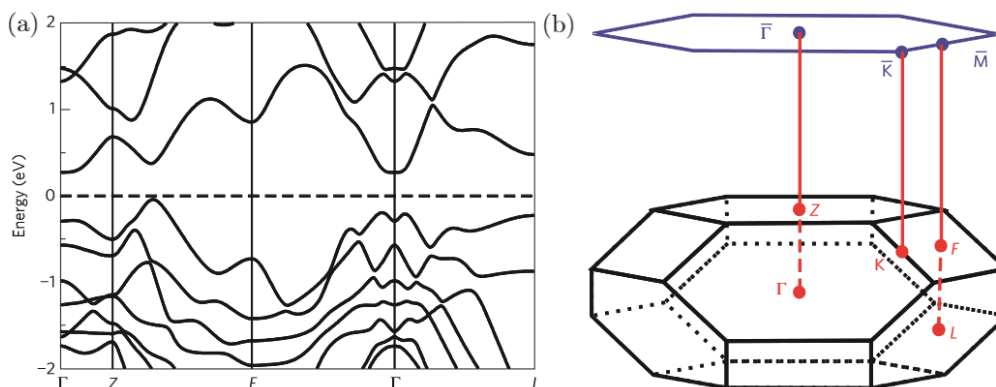
FIGURE 2: (a) Band structure of Bi_2Se_3. Here Γ, Z, F and L are convential names for points of high symmetry in the Brillouin zone, whose coordinates are $(0,0,0)$, $\left(\frac{1}{2}, \frac{1}{2}, \frac{1}{2}\right)$, $\left(\frac{1}{2}, \frac{1}{2}, 0\right)$ and $\left(\frac{1}{2}, 0, 0\right)$ respectively, in the dual basis $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$. The Brillouin zone and the positions of these high-symmetry points is shown in (b). The band structure is shifted so that the Fermi level lies at 0 eV. This material has a band gap. From Zhang et al. [15].

Band structures hold the key to determining the basis electronic conduction properties of materials, such as whether a material is an insulator, a conductor or a semi-conductor. This can be qualitatively understood in a picture of non-interacting electrons as follows.

---

[2]We remark that in solid state physics, the Brillouin zone is constructed in a slightly different but equivalent way. This more traditional Brillouin zone is formed by constructing the so-called Wigner-Seitz cell of the reciprocal lattice B^*. We will use this formulation later.

If one considers the electrons in the material to be added one by one, each electron will occupy the eigenstate with the lowest available energy. Because electrons are fermions, each single-particle state admits only one electron. As soon as all electrons have been added, the highest energy among the occupied eigenstates is called the Fermi level ε_{F}. Suppose that the Fermi level lies inside a so-called band gap, which is an interval of energies Δ with

$$\varepsilon_n(\mathbf{k}) \notin \Delta$$

for any n and $\mathbf{k}$. Such a band gap separates occupied bands from from unoccupied bands. In this case there are no low-energy excitations, as it requires a threshold of energy to promote electrons into conducting states. Such materials are called insulators. If there is no band gap, the material is a conductor, and if the band gap is sufficiently small, one calls the material a semi-conductor. This formalism is called band theory. We remark that although band theory successfully captures the basic properties of the electronic structure of solids, it is far from the full story, as band theory assumes that there are no interactions between electrons. Fortunately, as we will see in later sections, the topological phases of matter that we are concerning ourselves with in this thesis can also be understood from the viewpoint of band theory [16].

2.2 The Tight-Binding Method

Here we introduce an important approach to obtain the band structure of solids, known as the tight-binding method. In the exposition that we give here we follow Ashcroft and Mermin [5]. In tight-binding models we view the atoms that constitute a crystal as weakly interacting. As an extreme case, we can think of the atoms in the crystal having a separation that is much larger than the spatial extent of the relevant orbital wavefunctions of the individual atoms, so that the electronic eigenstates are localized at the crystal sites and states localized at different sites have essentially zero overlap. The perspective then changes, and quantum states become complex linear combinations of atomic orbitals localized on the atomic sites of the crystal, rather than wavefunctions defined on $\mathbb{R}^d$ or $\mathbb{T}^d$. In the tight-binding method it is then assumed that the separation of the atoms in the crystal is such that the Hamiltonian only couples orbitals on atomic sites that are close to each other, so that other couplings can be neglected.

More generally, the Hilbert space of a tight-binding model is of the form[3] $\ell^2(B) \otimes \mathbb{C}^N$, where B is a Bravais lattice and we view $\mathbb{C}^N$ as a Hilbert space encoding any internal structure of the Bravais lattice sites. These internal degrees of freedom describe the different atoms in the unit cell, the orbitals of each atom, and spin. The matrix elements of a tight-binding Hamiltonian arise from the coupling of orbitals on atomic sites near each other. One typically assumes that only the nearest-neighbour matrix elements are non-zero. Written in the notation of second quantization as in equation (31), the terms of the Hamiltonian have the interpretation of the electrons hopping from one lattice site to the other. For this reason, these matrix elements are called hopping amplitudes. The hopping amplitudes can be obtained from first principles using the knowledge of the orbitals of the atoms under consideration.

---

[3]$\ell^2(B)$ is the space of complex square-summable sequences on B, or equivalently the space of complex square-integrable functions on B with respect to the counting measure, up to functions that integrate to zero.

A generic tight-binding Hamiltonian has the form

$$H_{\text{TB}} = \sum_{\langle i,j \rangle} t_{ij} c_i^\dagger c_j \tag{3}$$

where i and j are indices encoding the lattice site as well as one of the basis elements of the internal Hilbert space $\mathbb{C}^N$. It assumed that the internal space comes with a given orthonormal basis. The notation $\langle i,j \rangle$ means that the summation runs over all neighbouring pairs of Bravais lattice sites, where the precise definition of neighbouring pairs depends on the model at hand. Equation (3) is written in the notation of second quantization, and as we discuss in appendix A this expression is equivalent to

$$H_{\text{TB}} = \sum_{\langle i,j \rangle} t_{ij} |i\rangle \langle j| \tag{4}$$

for a single-particle system. We will use both notations in what follows, because depending on the context one of the notations can be preferred over the other.

For the Hilbert space $\ell^2(B)$ we denote the standard orthonormal basis by

$$\{ |\mathbf{m}\rangle : \mathbf{m} \in B \},$$

where the $|\mathbf{m}\rangle$ are normalized states supported on the Bravais lattice site $\mathbf{m}$. In the lattice version of Bloch's theorem we will need a discrete version of plane waves. For tight-binding models on a finite lattice with $\mathcal{N}$ unit cells we introduce discrete plane waves, which are states in $\ell^2(B)$ of the form

$$|\mathbf{k}\rangle = \frac{1}{\sqrt{\mathcal{N}}} \sum_{\mathbf{m} \in B} e^{i\mathbf{k}\cdot\mathbf{m}} |\mathbf{m}\rangle,$$

where $\mathbf{k}$ lies in the Brillouin zone $\mathfrak{B}$. We are now in a position to introduce the lattice version of Bloch's theorem, whose formulation is slightly different than that of the more traditional continuous version of the theorem.

Theorem [Bloch, lattice version]. Consider a quantum system modelled on the Hilbert space $\ell^2(B) \otimes \mathbb{C}^N$ where B is a Bravais lattice, with a Hamiltonian H satisfying the translational invariance condition

$$T_\mathbf{R} H T_{-\mathbf{R}} = H$$

for any $\mathbf{R}$ in B. Then H and each $T_\mathbf{R}$ can be simultaneously diagonalized, and the common eigenstates $|\psi\rangle$ can be chosen to be Bloch waves, which are states of the form

$$|\psi_n(\mathbf{k})\rangle = |\mathbf{k}\rangle \otimes |u_n(\mathbf{k})\rangle, \tag{5}$$

where $\mathbf{k} \in \mathfrak{B}$, $|\mathbf{k}\rangle \in \ell^2(B)$ and $|u_n(\mathbf{k})\rangle \in \mathbb{C}^N$.

An important remark regarding the Brillouin zone that also applies to the previous subsection is that for a finite lattice with periodic boundary conditions, not every value of $\mathbf{k} \in \mathfrak{B}$ is allowed. If $\mathbf{v}_1, \ldots, \mathbf{v}_d$ denote the primitive lattice vectors, then Bloch's theorem together with the fact that $N_i \mathbf{w}_i \equiv 0$ implies that

$$|\mathbf{k}\rangle = T_{N_i \mathbf{v}_i} |\mathbf{k}\rangle = e^{i\mathbf{k} \cdot N_i \mathbf{v}_i} |\mathbf{k}\rangle \implies \mathbf{k} \cdot N_i \mathbf{v}_i \in 2\pi\mathbb{Z}, \text{ for each } i = 1, \ldots, d.$$

Physically, this means that the plane waves have to match the periodicity of the lattice. We will call all $\mathbf{k}$ for which the above holds the discrete Brillouin zone, which is thus given by

$$\mathfrak{B}' = \left\{ \frac{n_1}{N_1} \mathbf{w}_1 + \ldots + \frac{n_d}{N_d} \mathbf{w}_d \ : \ n_i = 1, \ldots N_i \right\}.$$

Hence, there are $\mathcal{N}$ plane waves for a crystal with $\mathcal{N}$ unit cells. From the identity

$$\sum_{m=1}^{\mathcal{N}} e^{2\pi i(k-k')m/\mathcal{N}} = \mathcal{N}\delta_{kk'}, \quad k, k' \in \mathbb{Z}$$

it follows that

$$\langle \mathbf{k}'|\mathbf{k}\rangle = \frac{1}{\mathcal{N}} \sum_{\mathbf{m}' \in B} \sum_{\mathbf{m} \in B} e^{-i\mathbf{k}' \cdot \mathbf{m}' + i\mathbf{k} \cdot \mathbf{m}} \langle \mathbf{m}'|\mathbf{m}\rangle = \frac{1}{\mathcal{N}} \sum_{\mathbf{m} \in B} e^{i(\mathbf{k}-\mathbf{k}') \cdot \mathbf{m}} = \delta_{\mathbf{k}\mathbf{k}'}.$$

Hence, the plane waves $|\mathbf{k}\rangle$ for $\mathbf{k} \in \mathfrak{B}'$ form an orthonormal basis of $\ell^2(B)$.

The states $|u_n(\mathbf{k})\rangle$ appearing in Bloch's theorem are now elements of the N-dimensional internal Hilbert space $\mathbb{C}^N$, and they satisfy

$$H(\mathbf{k}) |u_n(\mathbf{k})\rangle = \varepsilon_n(\mathbf{k}) |u_n(\mathbf{k})\rangle,$$

where $\varepsilon_n(\mathbf{k})$ are the energy eigenvalues and $H(\mathbf{k})$ is the Bloch Hamiltonian, which for lattice models takes the form [22]

$$H(\mathbf{k}) = \langle \mathbf{k}| H |\mathbf{k}\rangle.$$

The total Hamiltonian can be reconstructed from the Bloch Hamiltonian via

$$H = \sum_{\mathbf{k} \in \mathfrak{B}'} |\mathbf{k}\rangle \langle \mathbf{k}| \otimes H(\mathbf{k}). \tag{6}$$

We now make a small digression into some of the mathematical ideas related to the constructions of this section. Although the discrete Brillouin zone $\mathfrak{B}'$ has finitely many $\mathbf{k}-$points, one recovers the full Brillouin zone $\mathfrak{B}$ in the limit of an infinite crystal. In this way, one can view the system as consisting of an ensemble of Hamiltonians and Hilbert spaces $\left\{ \left(H(\mathbf{k}), \mathbb{C}^N \right) \right\}_{\mathbf{k} \in \mathfrak{B}}$ labelled by points $\mathbf{k}$ on a torus. Mathematically, this structure is a vector bundle whose base space is the smooth manifold $\mathfrak{B}$ and whose fibers are copies of $\mathbb{C}^N$. This vector bundle is trivial, because it can be simply written as $\mathfrak{B} \times \mathbb{C}^N$. Therefore, this bundle has no interesting topological properties. However, each $H(\mathbf{k})$ has its own set of eigenstates $\{|u_n(\mathbf{k})\rangle\}_{n=1}^N$, of which those with an energy eigenvalue below the Fermi level ε_F are occupied by the electrons. The number of occupied states N_F does not depend on $\mathbf{k}$ if there is a band gap. We denote the occupied states by $\{|u_n(\mathbf{k})\rangle\}_{n=1}^{N_F}$. The vector subbundle of $\mathfrak{B} \times \mathbb{C}^N$ whose fibers are the N_F-dimensional Hilbert spaces spanned by the occupied states $\{|u_n(\mathbf{k})\rangle\}_{n=1}^{N_F}$ is, in general, a non-trivial bundle, meaning that it cannot be expressed as a product of two spaces. Intuitively, this means that the fibers of this bundle are twisted. This bundle is sometimes called the Bloch bundle, and from a mathematical point of view, the topology of this bundle gives rise to non-trivial topological phases of matter. The $\mathbb{Z}_2$-index, to be discussed in section 3, is a topological invariant of this bundle.

2.3 Symmetries in Tight-Binding Models

Here we review how symmetries are represented in quantum mechanics, and in particular symmetries of systems modelled on a lattice such as tight-binding models. The symmetries that we treat here are fundamental for the type of topological insulator that we will consider in later sections. In this treatment we follow [22]. An important result regarding symmetry in quantum mechanics is Wigner's theorem, which states that symmetries of a quantum mechanical system are represented by unitary or antiunitary operators on the underlying Hilbert space. For our purposes we will need symmetries of both types.

Recall that a unitary operator $U : \mathcal{H} \to \mathcal{H}$ satisfies

$$\langle \psi | U^\dagger U | \phi \rangle = \langle \psi | \phi \rangle \quad \text{for all } |\psi\rangle, |\phi\rangle \in \mathcal{H},$$

whereas an antiunitary operator $A : \mathcal{H} \to \mathcal{H}$ is characterized by

$$\langle \psi | A^\dagger A | \phi \rangle = \langle \psi | \phi \rangle^* \quad \text{for all } |\psi\rangle, |\phi\rangle \in \mathcal{H}.$$

The first symmetry that we discuss is inversion symmetry, also known as parity symmetry. For systems modelled on $\mathbb{R}^d$, inversion about the origin is defined by the map

$$\mathbb{R}^d \to \mathbb{R}^d$$
$$\mathbf{r} \mapsto -\mathbf{r}.$$

The restriction of this operation to a Bravais lattice gives the inversion operation for a lattice model. It is represented by a unitary operator $P : \mathcal{H} \to \mathcal{H}$ which is involutive, so that $P^2 = 1$. With P defined in this way, a Hamiltonian $H : \mathcal{H} \to \mathcal{H}$ is said to be inversion-symmetric if

$$PHP^{-1} = H. \tag{7}$$

When working with tight-binding models, one usually works with vectors of the form $|\mathbf{k}\rangle \otimes |u\rangle$, where $|u\rangle$ is an element of the internal Hilbert space $\mathbb{C}^N$. The action of P on such a state should send $|\mathbf{k}\rangle$ to $|-\mathbf{k}\rangle$, but the action on the internal Hilbert space may be non-trivial depending on what the internal structure is. In general, one writes

$$P |\mathbf{k}\rangle \otimes |u\rangle = |-\mathbf{k}\rangle \otimes \pi |u\rangle,$$

where π is a unitary operator acting on $\mathbb{C}^N$. If there are different atoms in the unit cell, it may be the case that π permutes the sites of different atoms. If orbitals with non-zero angular momentum are included, then these are affected as well. However, the spin degree of freedom is always unaffected by inversion since spin is an intrinsic property without reference to real space. In most cases, the presence of inversion symmetry depends only on the symmetry properties of the crystal. For Bloch Hamiltonians, the inversion-symmetry condition of equation (7) takes the form

$$\pi H(\mathbf{k})\pi = H(-\mathbf{k}).$$

An important remark to which we will refer later is that the inversion of the Brillouin zone in d spatial dimensions $\mathbf{k} \mapsto -\mathbf{k}$ has 2^d fixed points, namely those points in the Brillouin zone for which every coordinate is either 0 or $\frac{1}{2}$ in the basis of reciprocal lattice vectors. These points are called the time-reversal invariant momenta, and they are denoted by

$\Gamma_i \in \mathfrak{B}$, where $i = 1, \ldots, 2^d$. These points satisfy[4] $\Gamma_i = -\Gamma_i$, and at these points the inversion-symmetry condition takes the form

$$\pi H(\Gamma_i)\pi = H(\Gamma_i),$$

from which one can see that $H(\Gamma_i)$ and π commute. It follows that eigenstates $|u_n(\Gamma_i)\rangle$ of the Hamiltonian at the time-reversal invariant momenta can be chosen to have a well-defined parity eigenvalue:

$$\pi |u_n(\Gamma_i)\rangle = \xi_n(\Gamma_i) |u_n(\Gamma_i)\rangle,$$

where the involutivity of π forces $\xi_n(\Gamma_i) = \pm 1$. Later we will see that the parity eigenvalues $\xi_n(\Gamma_i)$ at the time-reversal invariant momenta play a key role in determining whether a material is a topological insulator or not.

We now discuss time-reversal symmetry. In contrast to most symmetries in quantum mechanics, time-reversal is represented by a antiunitary operator. The action of time-reversal is to invert the arrow of time, meaning that quantities based on a temporal derivative such as momentum change their sign, whereas quantities such as position remain invariant. In the simple case where there is no internal structure present, time-reversal is represented by a complex conjugation operator K, which conjugates everything to its right. For example, if the Hilbert space consists of wavefunctions on $\mathbb{R}^d$, we have

$$K\psi(\mathbf{r}) = \overline{\psi(\mathbf{r})}K.$$

Note that $K^2 = 1$. The reason that complex conjugation represents time-reversal is the Schrödinger equation in real-space for a particle with no internal degrees of freedom,

$$i\hbar\partial_t\psi(\mathbf{r}, t) = H\psi(\mathbf{r}, t).$$

The conjugated wavefunction $\overline{\psi(\mathbf{r}, t)}$ satisfies the conjugated Schrödinger equation

$$-i\hbar\partial_t\overline{\psi(\mathbf{r}, t)} = \overline{H}\,\overline{\psi(\mathbf{r}, t)},$$

where $\overline{H} = KHK$. Since the left-hand side carrying the temporal derivative has changed sign after the conjugation, replacing operators and wavefunctions by their conjugate has the effect of time-reversal.

The usage of an operator of this type is quite subtle, because its definition depends on which basis is used. We define it on the real-space basis. In this way, K captures the properties that we expect of a time-reversal operator, since we have

$$Kx_jK^{-1} = x_j, \quad Kp_jK^{-1} = -p_j,$$

where x_j and p_j are the position and momentum operator in the jth direction, respectively. The latter equation follows from the fact that the momentum operator p_j is represented by $-i\partial_j$ in the real-space basis. If the particle that we describe has an internal structure such as spin, the definition of the time-reversal operator has to be extended to the internal Hilbert space. For our discussion of topological insulators, we will be interested in electrons, which have spin-$\frac{1}{2}$. In this case, the internal Hilbert space is $\mathrm{Sp}_{\mathbb{C}}\{|\uparrow\rangle, |\downarrow\rangle\} \cong \mathbb{C}^2$, where

---

[4]This equality is to be understood as an equality of coordinates on the torus, in the same way that $e^{i\pi} = e^{-i\pi}$ on the circle. Equivalently, it can be understood to hold modulo B^*.

$\text{Sp}_{\mathbb{C}}$ denotes the complex span and $|\uparrow\rangle$ and $|\downarrow\rangle$ denote the conventional spin eigenstates along the z-direction, which we identify with the column vectors

$$\begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix},$$

respectively. Since spin is an intrinsic form of angular momentum, it should change sign under time-reversal. Hence, we require a time-reversal operator τ acting on the internal Hilbert space that changes the sign of the spin matrices, so

$$\tau \sigma_j \tau^{-1} = -\sigma_j$$

for $j = x, y, z$. An operator that satisfies this condition is

$$\tau = \exp(i\pi\sigma_y/2)K = -i\sigma_y K, \tag{8}$$

which has the matrix representation

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} K,$$

and usually one chooses this operator to represent time-reversal on the spin-degree of freedom. Note that this operator also has the property $\tau^2 = -1$. At first glance this might seem incorrect for a time-reversal operator, as one may expect that inverting the arrow of time twice should leave a system invariant. However, the fact that the time-reversal operator squares to -1 is in fact a fundamental property of fermions. The reason is that spinors, the mathematical objects describing spin, behave non-trivially when rotated by 2π-rotation: their sign changes. The operator defined in equation (8) is precisely a rotation by an angle of π in the space of spinors, so that the square of the time-reversal operator corresponds to a 2π-rotation.

The time-reversal operator for the total system is then taken to be $\Theta = (1 \otimes -i\sigma_y)K$, where 1 is the identity on the space of wavefunctions. It inherits the fundamental property that $\Theta^2 = -1$. It is customary to denote this operator simply by $\Theta = -i\sigma_y K$, where it is implicitly understood that σ_y acts only on the spin degree of freedom of the electrons. As with the parity operator, a Hamiltonian H is said to be time-reversal symmetric if

$$\Theta H \Theta^{-1} = H \tag{9}$$

for an appropriate time-reversal operator Θ, usually based on the one mentioned above. For a tight-binding model on a finite periodic lattice with plane waves of the form

$$|\mathbf{k}\rangle = \frac{1}{\sqrt{\mathcal{N}}} \sum_{\mathbf{m} \in B} e^{i\mathbf{k}\cdot\mathbf{m}} |\mathbf{m}\rangle$$

we see that complex conjugation yields $K|\mathbf{k}\rangle = |-\mathbf{k}\rangle$. Therefore, time-reversal symmetry of the total Hamiltonian implies that

$$H = \Theta H \Theta^{-1} = \sum_{\mathbf{k} \in \mathfrak{B}'} |-\mathbf{k}\rangle \langle -\mathbf{k}| \otimes \tau \overline{H(\mathbf{k})} \tau^\dagger = \sum_{\mathbf{k} \in \mathfrak{B}'} |\mathbf{k}\rangle \langle \mathbf{k}| \otimes \tau \overline{H(-\mathbf{k})} \tau^\dagger,$$

and since

$$H = \sum_{\mathbf{k} \in \mathfrak{B}'} |\mathbf{k}\rangle \langle \mathbf{k}| \otimes H(\mathbf{k}),$$

it follows that the time-reversal symmetry condition for the Bloch Hamiltonian $H(\mathbf{k})$ is given by

$$H(\mathbf{k}) = \tau \overline{H(-\mathbf{k})} \tau^\dagger.$$

At the time-reversal invariant momenta Γ_i, this condition becomes

$$H(\Gamma_i) = \tau \overline{H(\Gamma_i)} \tau^\dagger,$$

which explains their name. This leads to an essential property of time-reversal operators that satisfy $\Theta^2 = -1$ called Kramers degeneracy, explained by a result called the Kramers theorem [22].

Theorem [Kramers]. Consider a quantum system with the same setup as above described by a time-reversal symmetric Hamiltonian H, with a time-reversal operator Θ satisfying $\Theta^2 = -1$. Then if $|\mathbf{k}\rangle \otimes |u_n(\mathbf{k})\rangle$ is an eigenstate of the Hamiltonian H, the time-reversed state

$$\Theta \, |\mathbf{k}\rangle \otimes |u_n(\mathbf{k})\rangle = |-\mathbf{k}\rangle \otimes \tau \, |u_n(\mathbf{k})\rangle$$

is also an eigenstate of H with the same energy eigenvalue, and this eigenstate is orthogonal to $|\mathbf{k}\rangle \otimes |u_n(\mathbf{k})\rangle$. Hence, each eigenstate of H is at least doubly degenerate.

The implications of the Kramers theorem for the time-reversal invariant momenta Γ_i are even stronger. Since time-reversal maps Γ_i onto Γ_i, each eigenstate $|u_n(\Gamma_i)\rangle$ of the Bloch Hamiltonian $H(\Gamma_i)$ is at least doubly denegerate in the internal Hilbert space $\mathbb{C}^N$. These pairs of eigenstates related by time-reversal are called Kramers pairs. As we will see in the next sections, this property is essential for topological insulators, and we will refer to this result many times.

2.4 Topological Equivalence of Lattice Hamiltonians

In the introduction we mentioned that one can assign topological invariants to quantum systems that cannot change if one continuously changes the Hamiltonian of the system, unless the system undergoes a topological phase transition. For this to make sense, we need to establish a notion of continuous deformation for a quantum system. In general, if we have a space $\mathscr{H}$ of Hamiltonians acting on a lattice system in which it makes sense to continuously deform a Hamiltonian[5], two Hamiltonians that both have a band gap at the Fermi level are said to be adiabatically connected if there exists a continuous path in $\mathscr{H}$ that links the two Hamiltonians, such that the band gap does not close on this path. In more physical words, two Hamiltonians are adiabatically connected if we can slowly change one into the other without closing the band gap. In the case that there is an important symmetry present, such as time-reversal symmetry, we restrict the notion of adiabatic continuity to include only deformations of the Hamiltonian that preserve this symmetry. In such cases, the topological properties of the Hamiltonian are said to be protected by that symmetry. For a given lattice model, this topological equivalence defines an equivalence relation on the corresponding set of gapped lattice Hamiltonians, and these equivalence classes can be assigned well-defined topological invariants. In this way, the topological invariants can only change if the system undergoes a topological phase transition, which necessarily implies that the band gap closes.

---

[5]Usually $\mathscr{H}$ is a subset of the space of bounded linear operators on the considered Hilbert space, which has a topology induced by the norm.

3   Time-Reversal Symmetric Topological Insulators

In this section we introduce the principles of topological insulators. We study a specific class of topological insulators, namely those that are symmetric under time-reversal. The first topological insulators that were discovered theoretically, called Chern insulators, were of a different type. Specifically, Chern insulators require that time-reversal symmetry is broken due to the presence of an external magnetic field or magnetic order. In contrast, time-reversal symmetric topological insulators can exist without external magnetic fields, making them more intrinsic. Instead, time-reversal symmetric topological insulators arise from spin-orbit coupling.

The time-reversal symmetric topological insulator is characterized by a topological invariant that takes values in $\mathbb{Z}_2$, which we will call the $\mathbb{Z}_2$-index. We begin this section by discussing the implications of this index for the phenomena at the surface of a topological insulator, and we describe the special properties that emerge. We then move back to the bulk, and review the origin of the $\mathbb{Z}_2$-index in terms of the bulk band structure due to Fu, Kane and Mele [7], [10]. After that we introduce two methods to compute the $\mathbb{Z}_2$-index, which are implemented numerically in section 4. We conclude by discussing a simple two-dimensional model to illustrate time-reversal symmetric topological insulators, called the Bernevig-Hughes-Zhang model.

3.1   Qualitative Description of the $\mathbb{Z}_2$-index

Following the basic classification of electronic phases into insulators and conductors of the previous section, a topological insulator belongs to the class of insulators, meaning that the bulk band structure has a band gap at the Fermi level separating the occupied bands from the conducting bands. A topological insulator distinguishes itself from an ordinary insulator by the following remarkable property: at the surface of the material the band gap closes, and gapless states emerge. In other words, a charge-carrying current can flow only on the surface of the material.

What makes these surface states special is that they are topologically protected, meaning that no adiabatic deformation of the Hamiltonian that respects time-reversal symmetry can destroy their existence. Another interesting aspect of these gapless surface states is that the dispersion near the points where the energy bands corresponding to surface states cross the Fermi level, called Dirac points, is linear, which means that the electrons at the surface can be phenomenologically described by the Dirac equation. The Dirac equation is characterized by a Hamiltonian with a linear dependence on momentum, and it is known for describing relativistic massless fermions. In other words, electrons in these states behave as if they have no mass. The spin of these gapless surface states is locked at a right-angle to their momentum, a phenomenon called spin-momentum locking. For this reason, surface states travelling in opposite directions have orthogonal spins, which strongly suppresses backscattering. In two-dimensional samples this has strong implications: the electrons in the edge states propagate around the sample essentially without reflection. For three-dimensional samples this results in a reduced resistivity. The spin-momentum locking also implies that the electrons do not only transport charge, but they also transport spin. This makes topological insulators interesting for the field of spintronics.

The topological nature of the surface or edge states is characterized by the number of surface or edge states there are present. We illustrate this for a two-dimensional time-reversal symmetric topological insulator with one direction of translational invariance and one direction of finite length, thus having two edges [22]. Figure 3 shows a possible band structure for such a setup. Edge states can be present in ordinary insulators, but adiabatic deformations can remove them. Specifically, an adiabatic deformation of the Hamiltonian that does not close the bulk band gap can only change the number of edge states at a given energy in the band gap in multiples of four, as demonstrated in figure 3. Meanwhile, due to Kramers degeneracy, the edge states come in pairs of two states. This means that if the number of Kramers pairs of edge states is odd, then under any adiabatic change of the Hamiltonian that preserves time-reversal symmetry there must always remain at least one pair of edge states. Hence, it is the parity of the number of pairs of edge states that characterizes the topological phase of the material.
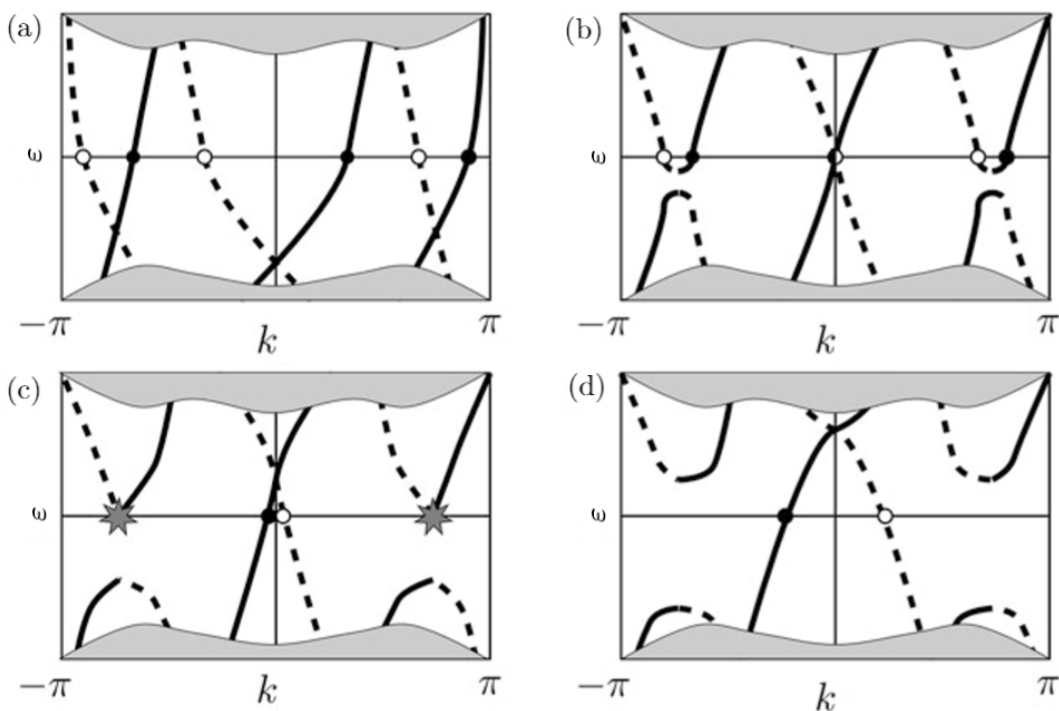
FIGURE 3: Band structure of a two-dimensional topological insulator with translational invariance along one direction and two edges, described by a wavenumber k in the one-dimensional Brillouin zone circle. Due to time-reversal symmetry, the spectrum is symmetric under $k \mapsto -k$. Continuous (dashed) lines show edge states travelling to the right (left). From (a) to (d) it is demonstrated that the number of edge states can only change in multiples of four. (a) Six edge state branches cross the Fermi level, corresponding to three Kramers pairs of edge states. (b) A small perturbation can turn the crossing edge state branches into avoided crossings. (c)-(d) The avoided crossings can be lifted above the Fermi level, reducing the number of edge states by four in the process. One pair of Kramers edge states remains, and this pair cannot be removed. Adapted from [22].

If $N(\varepsilon)$ is the number of edge states at a given energy ε in the bulk band gap, then we can set[6]

$$\nu = \frac{N(\varepsilon)}{2} \text{ mod } 2. \tag{10}$$

The number $\nu \in \mathbb{Z}_2$ is the $\mathbb{Z}_2$-index that we have been referring to. In view of the discussion of the preceding paragraph, a time-reversal symmetric topological insulator is characterized by $\nu = 1$. As a matter of fact, the $\mathbb{Z}_2$-index ν is purely defined in terms of a bulk Hamiltonian describing a system without edges, but it is the bulk-boundary correspondence that connects the bulk index ν with the parity of the number of pairs of edge states, hence making equation (10) valid.

This idea can be generalized to three dimensions [12]. In the three-dimensional case, the analogue of a one-dimensional branch crossing the Fermi level is a cone that lies inside the bulk band gap. This cone is called a Dirac cone, and in a three-dimensional topological insulator the Dirac cone is topologically protected by time-reversal symmetry. For a three-dimensional bulk system one can define four topological invariants ν_1, ν_2, ν_3 and ν. The ν_i are referred to as weak $\mathbb{Z}_2$-indices. The phase characterized by these weak indices is not robust against disorder, but the fourth index ν, called the strong index, characterizes a topological insulator in the sense that we have described above. In this case one could, through the bulk-boundary correspondence, interpret ν as the parity of the number of surface Dirac cones. The next subsection is devoted to the formulation of ν.

3.2   The $\mathbb{Z}_2$-index from the Bulk

In this subsection we give a brief overview of the formulation of the $\mathbb{Z}_2$-index ν in terms of the electronic structure of the bulk. The proper formulation is quite involved and lengthy, and hence we refer to [7], [10] for the details. As we mentioned earlier, the $\mathbb{Z}_2$-index arises from non-trivial topological properties of the Bloch bundle, which has the Brillouin zone torus $\mathfrak{B}$ as its base and fibers given by the subspaces of $\mathbb{C}^N$ spanned by the occupied eigenstates.

Following [10], we introduce the expression of the $\mathbb{Z}_2$-index in terms of so-called Wannier functions. Wannier functions are wavefunctions constructed from Bloch states that have the property of being localized in a chosen unit cell. We first consider a one-dimensional crystal $B = \mathbb{Z}_{\mathcal{N}}$ with $\mathcal{N}$ unit cells with periodic boundary conditions described by a Bloch Hamiltonian depending periodically on time t, satisfying

$$H(t) = H(t+T),$$
$$H(-t) = \Theta H(t) \Theta^{-1},$$

where T is the time period of the Hamiltonian and Θ is the time-reversal operator. We consider a system with N internal degrees of freedom and N_F occupied bands. In a cycle $t \in [0,T)$, there are two times, $t = 0$ and $t = T/2$, where the Hamiltonian is time-reversal symmetric. The eigenstates of the Hamiltonian are expressed in terms of Bloch states as

$$|\psi_n(k)\rangle = \frac{1}{\sqrt{\mathcal{N}}} e^{ikx} |u_n(k)\rangle,$$

---

[6]This number is ill-defined for energies at which the eigenvalue crossings are not simple, meaning that the derivative of the eigenvalue branch $\varepsilon(k)$ vanishes. The energies at which this occurs form a set of measure zero [22], and are thus ignored.

where k lies in the Brillouin zone $\mathfrak{B}$, which is a circle for this one-dimensional system. There is a certain freedom in how one chooses the eigenstates $|u_n(k)\rangle$ as a function of k, since states that differ by a phase factor describe the same state. Typically, one demands that the states $|u_n(k)\rangle$ vary smoothly as k is varied, and one calls this choice of states a *smooth gauge*. Given the states $|u_n(k)\rangle$, a Wannier function with band index $n \in \{1, \dots, N_{\mathrm{F}}\}$ centered at the unit cell positioned at $R \in B$ is defined by

$$|R, n\rangle = \frac{1}{2\pi} \oint_{\mathfrak{B}} dk\, e^{-ik(R-r)} |u_n(k)\rangle.$$

As we mentioned before, the state $|R, n\rangle$ is a superposition of Bloch states that has the property of being localized at R. These functions are not uniquely defined for a given Hamiltonian, but depend on the chosen gauge of the eigenstates $|u_n(k)\rangle$. In fact, in order to define Wannier functions the $|u_n(k)\rangle$ are not required to be eigenstates of the Bloch Hamiltonian, as long as they span the subspace of occupied eigenstates. In this case, a choice of gauge means a choice of unitary rotation of the eigenstates as k and t are varied. In this case one says that the gauge group is $U(N_{\mathrm{F}})$, the group of $N_{\mathrm{F}} \times N_{\mathrm{F}}$ unitary matrices. It can be shown that the *charge polarization P_ρ* of the crystal, defined in terms of the position expectation values of the Wannier functions of the occupied bands, can be expressed as

$$P_\rho = \sum_{n=1}^{N_{\mathrm{F}}} \langle 0, n| x |0, n\rangle = \frac{1}{2\pi} \oint_{\mathfrak{B}} dk\, \mathcal{A}(k), \tag{11}$$

where x is the position operator. Since the square of the absolute value of the Wannier functions represent a distribution of charge, the position expectation values $\langle 0, n| x |0, n\rangle =: \overline{x}_n$ are called the Wannier charge centers. The object $\mathcal{A}(k)$ is the so-called Berry connection,[7] defined by

$$\mathcal{A}(k) = i \sum_{n=1}^{N_{\mathrm{F}}} \langle u_n(k)| \partial_k |u_n(k)\rangle.$$

For a crystal with translational invariance, the charge polarization only makes sense up to a lattice vector, and hence the expression for P_ρ can be taken modulo multiples of the lattice constant. In this way, it is defined on a circle. If the Hamiltonian is time-reversal symmetric, the states $|u_n(k)\rangle$ for $n = 1, \dots, N$ can be grouped into Kramers pairs of states that are related by time-reversal, and written as $|u_n^{\mathrm{I}}(k)\rangle$, $|u_n^{\mathrm{II}}(k)\rangle$ for $n = 1, \dots, N/2$. Fu and Kane proposed to write this polarization as sum of two terms [10],

$$P_\rho = P^{\mathrm{I}} + P^{\mathrm{II}},$$

where the separate terms, called partial polarizations, are the charge polarizations corresponding to the states $|u_n^{\mathrm{I}}(k)\rangle$ and $|u_n^{\mathrm{II}}(k)\rangle$ respectively, defined by

$$P^{\mathrm{I}} = \frac{1}{2\pi} \oint_{\mathfrak{B}} dk\, \mathcal{A}^{\mathrm{I}}(k), \quad P^{\mathrm{II}} = \frac{1}{2\pi} \oint_{\mathfrak{B}} dk\, \mathcal{A}^{\mathrm{II}}(k),$$

---

[7]For the reader who is familiar with the terminology of differential geometry, the Berry connection is a connection in the sense of parallel transport on a principal fiber bundle. The corresponding Lie group in this case is $U(N_{\mathrm{F}})$, which encodes the unitary rotation of the states in the Bloch bundle as they are transported over the Brillouin zone.

where $\mathcal{A}^{\mathrm{I}}(k)$ and $\mathcal{A}^{\mathrm{II}}(k)$ are the Berry connections

$$\mathcal{A}^{\mathrm{I}}(k) = i \sum_{n=1}^{N_{\mathrm{F}}/2} \left\langle u_n^{\mathrm{I}}(k) \right| \partial_k \left| u_n^{\mathrm{I}}(k) \right\rangle, \quad \mathcal{A}^{\mathrm{II}}(k) = i \sum_{n=1}^{N_{\mathrm{F}}/2} \left\langle u_n^{\mathrm{II}}(k) \right| \partial_k \left| u_n^{\mathrm{II}}(k) \right\rangle.$$

Fu and Kane then proposed the concept of *time-reversal polarization*, denoted by P_θ and defined by the difference in the partial polarizations,

$$P_\theta = P^{\mathrm{I}} - P^{\mathrm{II}}.$$

The time-reversal polarization encodes the difference in charge polarization of the Kramers pairs of eigenstates. Since the time-dependent Bloch Hamiltonian $H(t)$ that we are considering has time-reversal symmetry at $t = 0$ and $T/2$ in a cycle $0 \le t < T$, there are two points in a cycle where the system has a well-defined time-reversal polarization. Under the assumption that the eigenstates $|u_n(k,t)\rangle$ evolve smoothly for $t \in [0, T/2]$, corresponding to a smooth choice of gauge, the difference in time-reversal polarization

$$\nu = P_\theta(T/2) - P_\theta(0) \bmod 2 \tag{12}$$

was shown to be a topological invariant of the Hamiltonian $H(t)$ with values in $\mathbb{Z}_2$. If a two-dimensional system is considered and (k, t) is replaced by (k_x, k_y) in the above discussion[8], then this ν is the definition of the $\mathbb{Z}_2$-index of a two-dimensional topological insulator. One way to think about this invariant is that the Wannier charge centers at $t = 0$ and $t = T/2$ come in pairs due to Kramers degeneracy, but during the half-cycle $t \in [0, T/2]$ they may flow over the circle and reconnect in a non-trivial way. For instance, they may switch partners in the sense that the Kramers pairs at $t = 0$ are no longer the same pairs as those at $t = T/2$. Using equation (12) as a starting point, Fu and Kane derived the expression

$$(-1)^\nu = \prod_i \frac{\sqrt{\det(w(\Gamma_i))}}{\mathrm{Pf}(w(\Gamma_i))}, \tag{13}$$

as a formula to compute the $\mathbb{Z}_2$-index, where the product runs over the two-dimensional time-reversal invariant momenta Γ_i, Pf is the Pfaffian of an antisymmetric matrix, and w is the matrix whose matrix elements are defined by

$$w_{mn}(\mathbf{k}) = \langle u_m(-\mathbf{k})| \Theta |u_n(\mathbf{k})\rangle,$$

where we assume a continuous gauge of the eigenstates $|u_n(\mathbf{k})\rangle$ for $\mathbf{k} \in \mathfrak{B}$. The Pfaffian of an antisymmetric matrix A satisfies

$$\left(\mathrm{Pf}(A)\right)^2 = \det(A),$$

and since $w(\mathbf{k})$ is antisymmetric at the Γ_i, equation (13) is well-defined. The fact that a continuous gauge is required in equation (13) makes it notoriously difficult to compute the $\mathbb{Z}_2$-index from this expression. Fortunately, simpler methods have been developed to compute it, which we will review in the next subsections.

---

[8]The replacement of time t by k_y and vice versa is common in discussions of this type, because it allows us to consider different interpretations of a process. It is called dimensional extension and dimensional reduction [22].

3.3 Fu-Kane Method - Parity at Time-Reversal Invariant Momenta

A while after Fu and Kane introduced equation (13) as a way to compute the $\mathbb{Z}_2$-index of a given system, they derived a simplification of equation (13) for crystals with inversion symmetry [11]. As we remarked in section 2, if the Bloch Hamiltonians $H(\mathbf{k})$ are invariant under inversion symmetry represented by an operator π, then the states $|u_n(\Gamma_i))\rangle$ at the time-reversal invariant momenta have a well-defined parity $\xi_n(\Gamma_i) \in \{-1, 1\}$. Due to presence of Kramers degeneracy, we can consider the N_{F} occupied states at Γ_i to consist of $N_{\mathrm{F}}/2$ Kramers pairs. Because parity and time-reversal commute, the two states in a Kramers pair have the same parity eigenvalue, so it makes sense to speak of the parity of a Kramers pair. Fu and Kane have shown that the factors in equation (13) involving the matrices $w(\mathbf{k})$ can be expressed in terms of the parity eigenvalues of Kramers pairs at the time-reversal invariant momenta, and hence that the $\mathbb{Z}_2$-index ν can be expressed as

$$(-1)^\nu = \prod_i \prod_n \xi_n(\Gamma_i), \tag{14}$$

where i runs over the 2^d time-reversal invariant momenta and n runs over the $N_{\mathrm{F}}/2$ Kramers pairs. This expression is very simple compared to the original expression, and it is straightforward to implement numerically, mainly because it does not require a smooth gauge over the full Brillouin zone. A smooth gauge is not required because the parity eigenvalues are gauge invariant. All one needs is a band structure with the corresponding eigenstates, which can be obtained using a tight-binding model, and the matrix representation of the inversion operator π. The simplicity of this method comes at the price of the restricted class of materials to which it can be applied, as it can only be applied to crystals with inversion symmetry.

3.4 Soluyanov-Vanderbilt Method - Wannier Charge Centers

An alternative method to compute the result of equation (12) was introduced by Soluyanov and Vanderbilt [18]. This method stays close to the original formulation in terms of Wannier charge centers, and in principle the method works for any crystal, without restrictions such as inversion symmetry. The major advantage of this method is that it does not require the manual construction of a smooth gauge over the whole Brillouin zone. Instead, a numerical procedure is used to automatically construct such a gauge. From this gauge, the Wannier charge centers $\bar{x}_n$ are computed using a discrete version of the Berry connection. We already remarked that the $\mathbb{Z}_2$-index can be thought of as a number that encodes the the way in which the Wannier charge centers flow over the circle as a function of t, and in particular whether they reconnect in a non-trivial way after half a cycle. By non-trivial we mean, for instance, that a pair of Wannier charge centers may split and flow in opposite directions over the circle, so that they reconnect with opposite winding numbers. The method to be discussed in this subsection is able to detect this non-trivial behaviour numerically. Some of the details of this method are complicated, so we refer to [18] for a detailed explanation.

In the Soluyanov-Vanderbilt method, the Wannier charge centers are obtained from an object called the Wilson loop. Let us assume that the states $|u_n(k, t)\rangle$ vary smoothly as one varies k over the Brillouin zone circle. Starting at a given k, we can traverse the Brillouin zone, varying each $|u_n(k, t)\rangle$ in the process, until we arrive at the same k again. The states $|u_n(k, t)\rangle$ at the end of the cycle are then not necessarily identical to the states

we started with. The Wilson loop is then defined to be the unitary rotation that relates the two sets of states at the start and at the end of a Brillouin zone cycle[9]. The first step of the numerical algorithm is to construct the Wilson loop. For the numerical implentation we have to discretize the Brillouin zone and the time period by choosing an appropriate number of gridpoints, which we denote by N_k and N_t respectively. For the discretization we set

$$k_i = \frac{i-1}{N_k - 1}, \ i = 1, \ldots, N_k, \qquad t_j = \frac{j-1}{N_t - 1}, \ i = 1, \ldots, N_t,$$

where we assume for the moment that the period of the Brillouin zone and the time cycle is 1 without loss of generality. One then diagonalizes the Bloch Hamiltonian on each (k_i, t_j) to obtain the eigenstates of the filled bands. This results in a discrete ensemble of eigenstates

$$\big\{ |u_n(k_i, t_j)\rangle : \ i = 1, \ldots, N_k, \ j = 1, \ldots, N_t, \ n = 1, \ldots, N_{\mathrm{F}} \big\},$$

where n is the band index and N_{F} is the number of occupied bands. At this point, the states of adjacent $k-$points are in principle unrelated. For this method, a smooth gauge is required for half of the time-cycle of the Hamiltonian. To obtain such a gauge numerically, we have to apply a $U(N_{\mathrm{F}})$-gauge transformation to the states at each gridpoint so that the states of adjacent gridpoints are as close together as possible. In [18] it is shown that the required gauge can be obtained by maximally localizing the Wannier functions, and that this gauge can be enforced numerically as follows. For each time t_j, one considers adjacent k-gridpoints and defines overlap matrices $M^{i,i+1}$ by

$$M_{mn}^{i,i+1} = \langle u_m(k_i, t_j) | u_n(k_{i+1}, t_j) \rangle .$$

Vanderbilt and Marzari have shown that the maximally localized Wannier function gauge requires each of the overlap matrices to be Hermitian [6]. For each i there is a unique gauge transformation that can be applied to the eigenstates at k_{i+1} to achieve this, which can be found by the singular value decomposition of $M^{i,i+1}$. We write

$$M^{i,i+1} = V \Sigma W^\dagger,$$

where V and W are unitary and Σ is a diagonal matrix containing the singular values of $M^{i,i+1}$. If one rotates the states as

$$|u_n(k_{i+1}, t_j)\rangle \mapsto W V^\dagger |u_n(k_{i+1}, t_j)\rangle ,$$

then the new overlap matrix $M^{i,i+1}$ is Hermitian. If this is done for each i, the result will be that the states at $k = 0$ and $k = 1$ are related by a unitary rotation Λ:

$$|u_n(k_1, t_j)\rangle = \Lambda |u_n(k_{N_k}, t_j)\rangle .$$

The matrix Λ is the Wilson loop, and it can be constructed by taking products of the overlap matrices. As we mentioned earlier, it is analogous to the integral of the Berry connection over a closed loop in the Brillouin zone. The eigenvalues $\lambda_{n,j} = e^{-2\pi i \overline{x}_{n,j}}$ of Λ are complex numbers of unit modulus, and it can be shown that the numbers $\overline{x}_{n,j} \in [-1/2, 1/2)$

---

[9]In the case that we are only changing one state as a function of k, the states at the endpoints of the cycle are related by a phase factor called the Berry phase. This Berry phase is the integral of the Berry connection along a closed loop, as in equation (11). For this reason, the Wilson loop is also called the non-abelian Berry phase, referring to the fact that the gauge group $U(N_{\mathrm{F}})$ is non-abelian.

are the Wannier charge centers at time t_j [22].

As we have seen, in [11] it was shown that the $\mathbb{Z}_2$-index is related to the motion of the Wannier charge centers during a half a time-cycle of the Hamiltonian. For a small number of bands and with a high time-resolution this motion can be inspected visually, but as soon as one deals with systems with many bands it can be difficult to decide when two Wannier charge center flows cross.

A numerically stable approach to extract this information was proposed in [18]. Instead of focusing on each Wannier charge center $\overline{x}_{n,j}$ individually, the largest gap between the charge centers is is tracked, denoted by z_j. If the time resolution is sufficiently high, the z_j take the form of a sequence of path segments with a number of discontinuities. The $\mathbb{Z}_2$-index is encoded in the number of Wannier charge centers that are crossed in these discontinuous jumps. It was shown that if Δ_j denotes the number of charge centers $\overline{x}_{n,j+1}$ between z_j and z_{j+1}, then the $\mathbb{Z}_2$-index is given by

$$\nu = \sum_j \Delta_j \bmod 2, \tag{15}$$

where the sum runs over the j corresponding to half a time-cycle. A simple and robust method to obtain the Δ_j proposed by Soluyanov and Vanderbilt is to consider the sign of the directed area of the triangle spanned by z_j, z_{j+1}, and $\overline{x}_{n,j+1}$. If ϕ_1, ϕ_2 and ϕ_3 are angles on the unit circle, then the directed area of the triangle defined by these angles can be expressed as

$$g(\phi_1, \phi_2, \phi_3) = \sin(\phi_2 - \phi_1) + \sin(\phi_3 - \phi_2) + \sin(\phi_1 - \phi_3).$$

In this way, it was shown that the Δ_j can be found by

$$(-1)^{\Delta_j} = \prod_{n=1}^{N_\mathrm{F}} \mathrm{sgn}\big(g(z_j, z_{j+1}, \overline{x}_{n,j+1})\big).$$

The steps outlined above can be turned into a computational scheme that turns a Bloch Hamiltonian directly into the $\mathbb{Z}_2$-index. In the next subsection we will apply this method to a simple model for a topological insulator to illustrate the features of the Wannier charge centers. Later in section 4 we will also apply the method to a realistic three-dimensional system. Since the discussion above assumes a two-dimensional systems, a generalization to three-dimensions is needed first. This is relatively straightforward and will be discussed in section 4.

3.5   The Bernevig-Hughes-Zhang Model

In this subsection we introduce a simple tight-binding model for a two-dimensional time-reversal symmetric topological insulator first introduced by Bernevig, Hughes and Zhang (BHZ) [9], [22]. The BHZ model originates from a theoretical study of the quantum spin Hall effect in HgTe quantum wells, which were the first experimentally realized topological insulators. This model has only four bands, which makes it an appropriate model to illustrate the basic principles of time-reversal symmetric topological insulators. After we have introduced it, we will compute its $\mathbb{Z}_2$-index for two different parameter values using the Soluyanov-Vanderbilt method. We will also investigate the associated edge states.

The BHZ model is defined on a two-dimensional square lattice with four states per unit cell. These states are a combination of two orbital states and two spin states. Choosing these four states as a basis, the BHZ model is defined [22] via the matrix Bloch Hamiltonian given by

$$H_{\text{BHZ}}(k_x, k_y) = s_0 \otimes \left[(u + \cos k_x + \cos k_y)\sigma_z + \sin k_y \sigma_y \right] + s_z \otimes \sin k_x \sigma_x + s_x \otimes C.$$

Here u is a real parameter that has the interpretation of an on-site potential, and C is a coupling operator between the two spinors. We set $C = 0.3\sigma_y$, as in [22]. The σ_j and s_j are Pauli matrices, where the σ_j act on the spin degree of freedom of and the s_j act on the orbital degree of freedom. The matrix s_0 denotes the 2×2 identity matrix. For this choice of coupling C, this model has time-reversal symmetry represented by the operator

$$\Theta = -i(s_0 \otimes \sigma_y)K,$$

which is of the same form as the one we encountered in section 2. The parameter u can be adjusted, and different topological phases are known to occur for different values. In particular, it is known that the band gap of the BHZ model closes near $u = 2$, $u = -2$ and $u = 0$, indicating possible topological phase transitions. We study the BHZ model at $u = -1.2$ and $u = -2.8$, so the corresponding Hamiltonians are not adiabatically connected due to the closure of the band gap near $u = -2$. Figure 4 shows the band structures for these two values, and figure 5 shows the band gap as a function of $u \in [-2.8, -1.2]$.

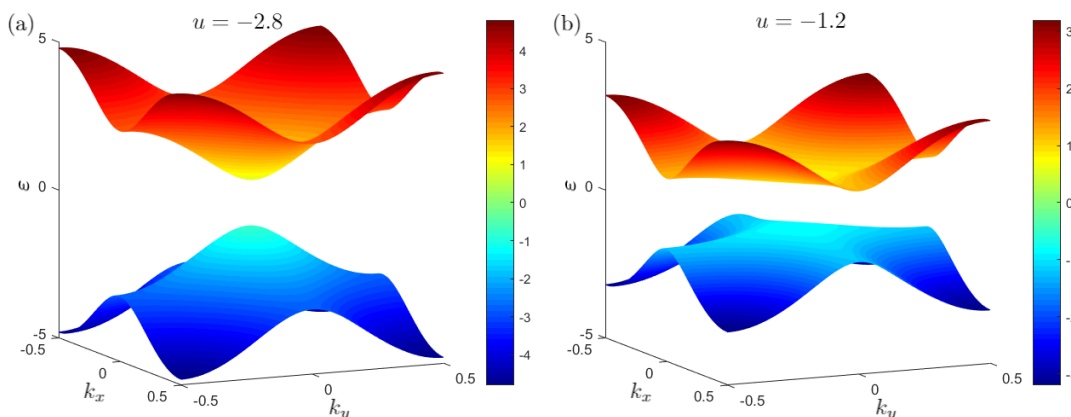FIGURE 4: Band structures of the BHZ model for different values of u, with $u = -1.2$ in (a) and $u = -2.8$ in (b). Both band structures have a band gap around $\varepsilon = 0$. Note that k_x, k_y and ε are dimensionless in the BHZ model.
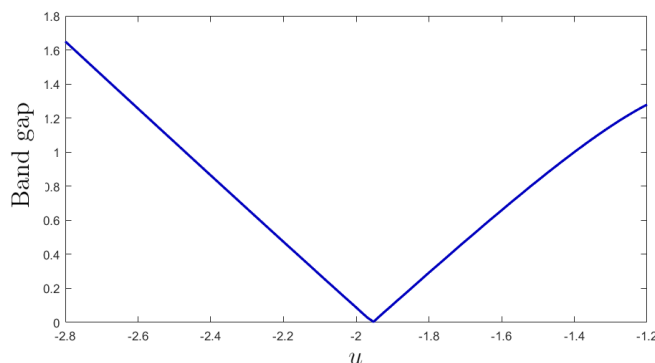

FIGURE 5: Band gap of the BHZ model for u between $u = -2.8$ and $u = -1.2$. The band gap closes near $u = -2$. This indicates a possible topological phase transition.

25

The BHZ model has time-reversal symmetry, so we can implement the Soluyanov-Vanderbilt method outlined in the previous subsection to compute the $\mathbb{Z}_2$-index. The computations have been done using MATLAB. Figure 6 shows the resulting Wannier charge center flows during one time-cycle of the Hamiltonian. For $u = -2.8$, one sees that that the Wannier charge centers slightly move as a function of t, but they reconnect in the same way as they started. In contrast, for $u = -1.2$ the charge centers separate and wind around the circle in opposite directions, reconnecting in a non-trivial way. In relation to equation (12), one can also see that for $u = -1.2$ the charge centers give rise to a different time-reversal polarization at $t = 0$ and $t = 1/2$. The computation returns that the $\mathbb{Z}_2$-index is $\nu = 0$ for the case $u = -2.8$ and $\nu = 1$ for the case $u = -1.2$. Therefore, these different choices of parameters correspond to different topological phases, separated by a topological phase transition accompanying the closure of the band gap. More specifically, the BHZ model with $u = -1.2$ is a time-reversal symmetric topological insulator, and therefore it should host topologically protected edge states. The edge states will be investigated in section 5.
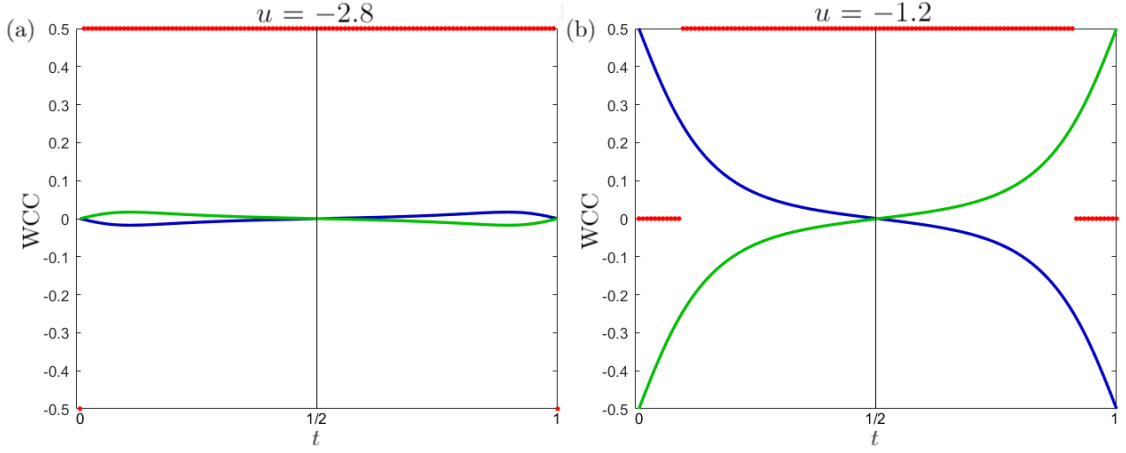
FIGURE 6: Wannier charge center (WCC) flow as a function of time during one time cycle for the BHZ model. The blue and green lines show the WCCs, and the red dots show the center of the largest gap between the WCCs (z_j). Note that the coordinates on the vertical axis are angular, so $0.5 \equiv -0.5$. (a) WCC flow for BHZ model with $u = -2.8$. The WCCs do not move non-trivially over the circle, and the numerical scheme returns $\nu = 0$ for the $\mathbb{Z}_2$-index, corresponding to a trivial insulator. (b) WCC flow for BHZ model with $u = -1.2$. The WCCs wind around the circle in opposite directions, and the numerical scheme returns $\nu = 1$ due to the discontinuous jump in the red line, corresponding to a topological insulator.

4 Topological Insulator in Bi$_2$Se$_3$

In this section we investigate the existence of a topological insulator in a real material, namely bismuth selenide (Bi$_2$Se$_3$). Both theoretically and experimentally, this material is known to be three-dimensional time-reversal symmetric topological insulators [14], [15]. The topologically non-trivial band structure of this material arises from relatively strong spin-orbit coupling, which causes a band inversion at the Γ point. In this section we present a computation of the band structures of Bi$_2$Se$_3$ compounds and we compute its $\mathbb{Z}_2$-index. Because the crystal structure of this material have inversion symmetry, this computation can be done using the Fu-Kane method discussed in the previous section. To demonstrate that this computation could have been done without the presence of inversion symmetry, we also apply the Soluyanov-Vanderbilt method to arrive at the same result. For the computation of the bulk band structure we use a Slater-Koster tight-binding method, following Pertsova and Canali [21] for the tight-binding Hamiltonian, Zhang et al. [15] for the crystal structure and Kobayashi [17] for the tight-binding parameters.

4.1 Crystal Structure of Bi$_2$Se$_3$

Bi$_2$Se$_3$ has a crystal structure with a rhombohedral unit cell containing five atoms, as shown in figure 7. The structure consists of five-atom layers of triangular lattice planes extending in the x- and y-direction, called quintuple layers, which are stacked in the z-direction.
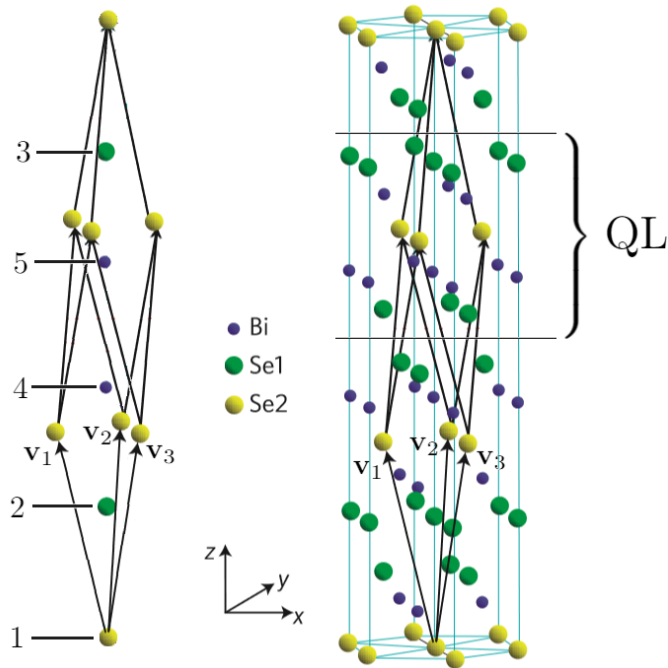
FIGURE 7: Crystal structure of Bi$_2$Se$_3$ with primitve lattice vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ spanning the Bravais lattice. Se atoms at inequivalent positions are labelled by Se1 and Se2. Left: rhombohedral unit cell, showing our convention for the labels of the atomic sites in the unit cell. Right: full crystal structure in the neighbourhood of a unit cell. A quintuple layer (QL) is indicated. Adapted from Zhang et al. [15].

Figure 7 also shows our convention for labelling the atoms in the unit cell. The crystal has inversion symmetry centered at any of the Se2 atoms, meaning that the crystal structure remains invariant under the parity operation $\mathbf{r} \mapsto -\mathbf{r}$ if one of the Se2 sites is placed at the origin. By looking at the crystal structure in figure 7, one sees that the parity

operation induces a permutation on the atomic sites in the unit cell given by

$$(1, 2, 3, 4, 5) \mapsto (1, 3, 2, 5, 4).$$

In other words, Se2 sites are mapped onto themselves, and the different Se1 and Bi sites in the unit cell switch places. We will refer to this permutation later when we construct the parity operator acting on the internal Hilbert space of our tight-binding model.

4.2   Slater-Koster Tight-Binding Hamiltonian

Following [21], we use a Slater-Koster tight-binding method with three p-orbitals [1]. This means that we assume that each of the five atoms in the unit cell admits three electronic orbital states, namely p_x, p_y and p_z. We include spin, so the electron has a spin degree of freedom which takes values in the two-dimensional Hilbert space spanned by $|\uparrow\rangle$ and $|\downarrow\rangle$, where the arrows $\uparrow$ and $\downarrow$ are conventional labels for the spin eigenstates along the z-direction. Combining these internal degrees of freedom, the internal Hilbert space of this system has the form

$$\mathcal{H} = \mathrm{Sp}_{\mathbb{C}}\big\{\, |1\rangle, \ldots, |5\rangle \,\big\} \otimes \mathrm{Sp}_{\mathbb{C}}\big\{\, |p_x\rangle, |p_y\rangle, |p_z\rangle \,\big\} \otimes \mathrm{Sp}_{\mathbb{C}}\big\{\, |\uparrow\rangle, |\downarrow\rangle \,\big\} \cong \mathbb{C}^5 \otimes \mathbb{C}^3 \otimes \mathbb{C}^2 \cong \mathbb{C}^{30}, \quad (16)$$

where $\mathrm{Sp}_{\mathbb{C}}$ denotes the complex span. Here $|1\rangle, \ldots, |5\rangle$ denote the eigenstates corresponding to the lattice sites in the unit cell according to the convention we introduced above. For $\mathbf{k}$ in the Brillouin zone $\mathfrak{B} \cong \mathbb{T}^3$, the Bloch Hamiltonian $H(\mathbf{k})$ acting on $\mathcal{H}$ then takes the form [21]

$$H(\mathbf{k}) = \sum_{ij,\sigma,\alpha\alpha'} t_{ij}^{\alpha\alpha'} e^{i\mathbf{k}\cdot\mathbf{r}_{ij}} c_{i\alpha}^{\sigma\,\dagger} c_{j\alpha'}^{\sigma} + \sum_{i,\sigma\sigma',\alpha\alpha'} \lambda_i \langle i, \alpha, \sigma | \mathbf{L} \cdot \mathbf{S} | i, \alpha', \sigma' \rangle c_{i\alpha}^{\sigma\,\dagger} c_{i\alpha'}^{\sigma'}. \quad (17)$$

Here i and j label the atomic sites, the indices α and α' denote the orbital of atom i and j respectively, and the spin of the electron states is denoted by σ and σ'. The index i runs over the five atoms in the unit cell, and j runs over all atoms in the neighbourhood of i, including atoms in adjacent unit cells. As in [17], we assume that there are only interactions between atoms in the same layer, in the nearest layer, and in the second nearest layer. This means that for each i, the index j runs over eighteen different atomic sites. The vector $\mathbf{r}_{ij}$ denotes the relative vector between the unit cells of atoms i and j. Finally, $c_{i\alpha}^{\sigma\,\dagger}$ and $c_{i\alpha}^{\sigma}$ denote the creation and annihilation operators of electrons with spin σ at the atomic site i in the orbital α. The parameters $t_{ij}^{\alpha\alpha'}$ are the hopping amplitudes, and the λ_i is the strength of the on-site spin-orbit coupling at atom i. The operators $\mathbf{L}$ and $\mathbf{S}$ are the orbital angular momentum and spin operators, respectively. The values of $t_{ij}^{\alpha\alpha'}$ can be deduced from the associated Slater-Koster parameters, which are a basic set of parameters from which the hopping amplitudes can be constructed using the relative orientations of the atoms in the unit cell. These have been calculated by Kobayashi [17] by fitting tight-binding band structures to band structures obtained using density functional theory. The conversion of the parameters in [17] is discussed in appendix B. If we choose a basis, we can use equation (17) to obtain the matrix-valued map

$$H : \mathfrak{B} \to \mathrm{Mat}_N(\mathbb{C})$$
$$\mathbf{k} \mapsto H(\mathbf{k}),$$

where $\mathrm{Mat}_N(\mathbb{C})$ is the algebra of N by N complex matrices. For a given $\mathbf{k} \in \mathfrak{B}$, the spectrum of $H(\mathbf{k})$ can be computed numerically, from which we can extract the bulk band

structure of Bi_2Se_3. The details of the passage from the tight-binding Hamiltonian to the tight-binding matrix are shown in appendix B. We have used MATLAB for the computations. Figure 8 shows the computed band structure of Bi_2Se_3. Due to the relatively strong spin-orbit coupling, the top valence band and bottom conduction band are inverted[10] at the Γ point. The band structure computed with this tight-binding model agrees with that of the literature, shown in figure 2, except for the band gap energy, which is slightly smaller in our band structure.
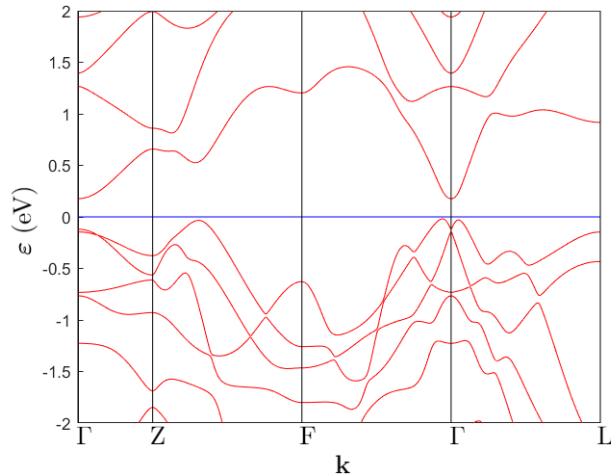

FIGURE 8: Computed band structure of Bi_2Se_3. The blue line indicates the Fermi level. A band inversion can be observed at the Γ point. The Brillouin zone is the same as the one shown in figure 2.

4.3 $\mathbb{Z}_2$-index of Bi_2Se_3 from the Fu-Kane Method

In this subsection we compute the $\mathbb{Z}_2$-index using the method proposed by Fu and Kane [11] for topological insulators with inversion symmetry. The spatial inversion symmetry that we mentioned earlier is represented by a unitary operator $\pi : \mathcal{H} \to \mathcal{H}$. This operator can be factorized as a tensor product $\pi = \pi_A \otimes \pi_O \otimes \pi_S$, where each factor acts on one of the tensor factors of the internal Hilbert space in (16). The subscripts A, O, and S stand for atom, orbit and spin respectively. As we have seen from the geometry of the unit cell, inversion has the effect of permuting the sites of the atoms, transposing the pairs (2,3) and (4,5). Hence, we can write

$$\pi_A = |1\rangle \langle 1| + |2\rangle \langle 3| + |3\rangle \langle 2| + |4\rangle \langle 5| + |5\rangle \langle 4| . \tag{18}$$

The p_x, p_y and p_z orbitals change sign under inversion, so we have

$$\pi_O = - |p_x\rangle \langle p_x| - |p_y\rangle \langle p_y| - |p_z\rangle \langle p_z| . \tag{19}$$

Finally, we recall that spin remains invariant under inversion. Therefore, π_S is simply the identity, or

$$\pi_S = |\uparrow\rangle \langle \uparrow| + |\downarrow\rangle \langle \downarrow| . \tag{20}$$

---

[10]A band inversion occurs when energy bands are pushed into each other by effects such as spin-orbit coupling. Near the avoided crossing, the eigenstates associated to the two bands switch in a certain sense. In 8 this can be seen from the small valley at the top of the valence band at the Γ point.

Taking the tensor product of equations (18), (19), and (20) gives us the full inversion operator π, where we remark that $|i, \alpha, \sigma\rangle = |i\rangle \otimes |\alpha\rangle \otimes |\sigma\rangle$, and

$$c_{i\alpha}^{\sigma\dagger} c_{j\alpha'}^{\sigma'} \equiv |i, \alpha, \sigma\rangle \langle j, \alpha', \sigma'|.$$

Observe that $\pi^2 = 1$ and that π is self-adjoint, as should be the case for an inversion operator. We remark, as a minor detail, that there are two different conventions for the phase factors $e^{i\mathbf{k}\cdot\mathbf{r}_{ij}}$ in equation (17), and depending on this convention the inversion operator may need to be adjusted. In the convention that we have chosen, the inversion operator must be conjugated by a diagonal unitary matrix $D(\mathbf{k})$ with the phase factors $e^{i\mathbf{k}\cdot\mathbf{r}_j}$ on the diagonal, with j the corresponding label of the atomic site, so

$$\pi \mapsto D(\mathbf{k})\pi D(\mathbf{k})^{-1}.$$

In particular, the inversion operator becomes $\mathbf{k}$-dependent.

From the eigenstates of the Hamiltonian at the time-reversal invariant momenta Γ_i, the parity eigenvalues $\xi_n(\Gamma_i) \in \{\pm 1\}$ can be computed. Table 1 shows products of the the parities of the occupied Kramers pairs at each time-reversal invariant momentum Γ_i.

TABLE 1: Products of the parity eigenvalues of the occupied Kramers pairs at the time-reversal invariant momenta $\Gamma_1, \ldots, \Gamma_8$. The Γ_i are expressed as $w_1\mathbf{w}_1 + w_2\mathbf{w}_2 + w_3\mathbf{w}_3$, and the corresponding coordinates w_1, w_2, w_3 are shown, together with the conventional names of the Γ_i.

	Γ_1 (Γ)	Γ_2 (L)	Γ_3 (L)	Γ_4 (L)	Γ_5 (F)	Γ_6 (F)	Γ_7 (F)	Γ_8 (Z)
w_1	0	1/2	0	0	1/2	1/2	0	1/2
w_2	0	0	1/2	0	1/2	0	1/2	1/2
w_3	0	0	0	1/2	0	1/2	1/2	1/2
$\prod_n \xi_n(\Gamma_i)$	-1	1	1	1	1	1	1	1

Using (14), we compute the $\mathbb{Z}_2$-index of Bi_2Se_3 by taking the product of the numbers in the bottom row, and we find that

$$\prod_i \prod_n \xi_n(\Gamma_i) = -1 \implies \nu = 1.$$

Hence, the $\mathbb{Z}_2$-index is 1, and we infer from the bulk-boundary correspondence that Bi_2Se_3 is indeed a topological insulator. In section 5 we will investigate the surface states of Bi_2Se_3 to confirm the existence of topologically protected gapless surfaces states.

4.4 $\mathbb{Z}_2$-index of Bi_2Se_3 from the Soluyvanov-Vanderbilt Method

Although we have in principle already computed the $\mathbb{Z}_2$-index of Bi_2Se_3, we relied on a method that only works for a special class of crystalline solids. In this section we demonstrate that the same results can be obtained using the Soluyanov-Vanderbilt method, which works for any three-dimensional crystal. This will also illustrate that this method indeed works when many energy bands are present.

Our first step is to review the generalization this method to three dimensions [18]. The reciprocal lattice vectors $\mathbf{w}_1/2, \mathbf{w}_2/2, \mathbf{w}_3/2$ span a parallelepiped in momentum-space, whose vertices are the time-reversal invariant momenta Γ_i. This object has six faces,

and each of these faces can be interpreted as a two-dimensional periodic system if the Hamiltonian is regarded as a function of two momentum parameters. In particular, each of these faces can be assigned a $\mathbb{Z}_2$-index, which can be computed using the Soluyanov-Vanderbilt method. Consider any two opposite faces, so that each of the Γ_i is the vertex of precisely one of the chosen faces, and let the $\mathbb{Z}_2$-indices of these faces be denoted by ν_{I} and ν_{II}. Then the strong $\mathbb{Z}_2$-index ν is given by

$$\nu = \nu_{\mathrm{I}} + \nu_{\mathrm{II}} \bmod 2. \tag{21}$$

In other words, for ν to be non-trivial one of the faces should be trivial whereas the opposite face is non-trivial. We note that the result is independent of the chosen pair of opposite faces.

Hence, with $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3$ the primitive reciprocal lattice vectors of the chosen $\mathrm{Bi}_2\mathrm{Se}_3$ unit cell, we compute ν_{I} and ν_{II} for the faces

$$k\mathbf{w}_1 + t\mathbf{w}_2 \quad \text{and} \quad \frac{1}{2}\mathbf{w}_3 + k\mathbf{w}_1 + t\mathbf{w}_2 \quad \text{for } k, t \in [0, 1],$$

which we will refer to as face 1 and face 2 respectively. The results of the computation are shown in figure 9.
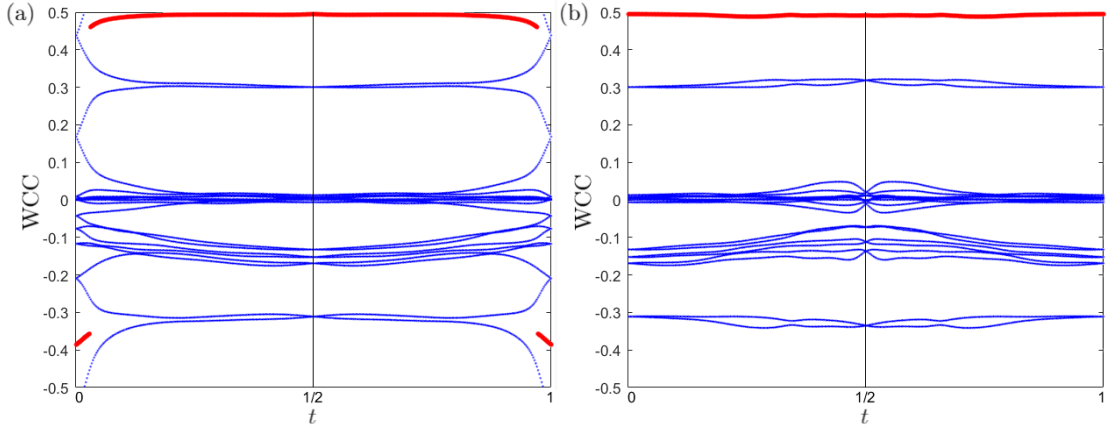

FIGURE 9: Wannier charge center (WCC) flow as a function of t during one cycle for $\mathrm{Bi}_2\mathrm{Se}_3$. Here the crystal momentum in the direction of $\mathbf{w}_2$ plays the role of t. The blue lines show the WCCs, and the red dots show the center of the largest gap between the WCCs. Note that the coordinates on the vertical axis are angular, so $0.5 \equiv -0.5$. (a) WCC flow at face 1. The WCCs separate and reconnect in a non-trivial way, switching partners after a cycle. The red line reflects this and shows a single discontinuity in the half-cycle $t \in [0, 1/2]$. The numerical scheme indeed returns $\nu_{\mathrm{I}} = 1$. (b) WCC flow at face 2. The WCCs move around but reconnect trivially, and the red line remains continuous. The numerical scheme returns $\nu_{\mathrm{II}} = 0$.

Figure 9 shows that one of the faces has a trivial $\mathbb{Z}_2$-index whereas the other one is non-trivial. The Wannier charge centers behave in a similar way as in figure 6. In particular, note that the time-reversal polarization of face 1 at $t = 0$ and $t = 1/2$ is different since the Wannier centers are positioned differently at these time points. Hence, the Soluyanov-Vanderbilt method also returns $\nu = 1$ through equation (21), in agreement with the result we obtained from the Fu-Kane method. This confirms that this method works for realistic

models of three-dimensional topological insulators. As a final remark we mention that one of the aims of the Soluyanov-Vanderbilt is to provide a robust computation of the $\mathbb{Z}_2$-index that does not rely on visual inspection of the Wannier charge centers. To make the flow of the charge centers visible we have used a relatively fine t-grid in figure 9, but the method gives the same result for coarser grids in which the flow of the charge centers is not directly apparent, showing that the method succeeds in that regard.

5   Surface States

In the previous sections we have seen a number of tight-binding models with non-zero $\mathbb{Z}_2$-indices, which indicates that these materials should be in a non-trivial topological phase and thus host topologically protected surface or edge states. The aim of this section is to detect and confirm the existence of these surface states. To this end, we will make use of the formalism of surface Green's functions, following [8] and [23]. With this formalism one can compute the local density of states at the surface, which can make the surface states visible. To keep the discussion general, we shall use the term boundary to refer to either a surface or an edge. We do not include every detail of the derivation of this method to keep the discussion concise.

5.1   Surface Green's Functions

The setup that we consider is a tight-binding model with N internal degrees of freedom on a semi-infinite lattice of the form $\widetilde{B} = \mathbb{N} \times B$, where B is a Bravais lattice which we take to be either just $\mathbb{Z}$ or $\mathbb{Z} \oplus \mathbb{Z}$. The difference with an infinite lattice is that there is one distinct axis in which the lattice only extends in one direction, similar to the geometry of a half-space. In this way, the system has a single boundary, corresponding to the layer $\{1\} \times B$. Note that we use the convention that $0 \notin \mathbb{N}$. An illustration of a setup of this type is given in figure 10.
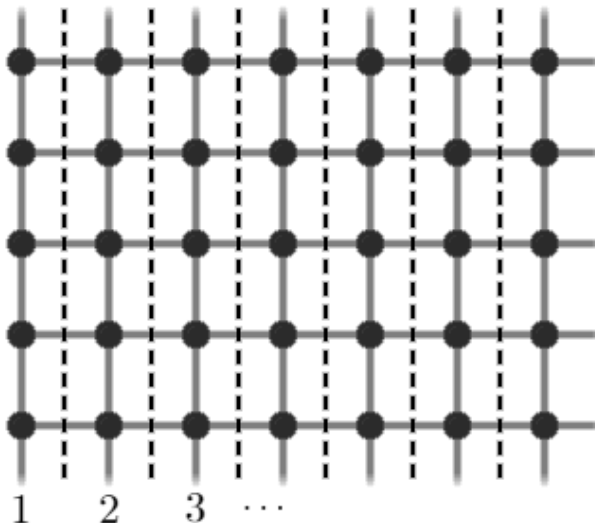
FIGURE 10: Two-dimensional semi-infinite square lattice $\mathbb{N} \times \mathbb{Z}$. The lattice is divided into layers parallel to the edge, indexed by a natural number i. The boundaries of the layers $\{i\} \times \mathbb{Z}$ are indicated by the dashed lines. The layer at $i = 1$ is the edge layer.

We subdivide the lattice into layers parallel to the boundary, indexed by $i \in \mathbb{N}$. This means that we decompose the Hilbert space $\mathcal{H} = \ell^2(\widetilde{B}) \otimes \mathbb{C}^N$ as a direct sum of layers,

$$\mathcal{H} = \bigoplus_{i=1}^{\infty} \mathcal{H}_i,$$

where $\mathcal{H}_i = \ell^2(\{i\} \times B) \otimes \mathbb{C}^N$ is the Hilbert space associated to the ith layer. On each layer we have translational symmetry parallel to the boundary, so we can use Bloch's theorem to label the eigenstates on each layer by a wavevector $\mathbf{k}$ parallel to the boundary. In the

two-dimensional case this is a scalar k_y, and in the three-dimensional case $\mathbf{k} = (k_x, k_y)$ is a vector. We subdivide the total Bloch Hamiltonian $H(\mathbf{k}) : \ell^2(\mathbb{N}) \otimes \mathbb{C}^N \to \ell^2(\mathbb{N}) \otimes \mathbb{C}^N$ into blocks $H_{i,j}(\mathbf{k}) : \mathbb{C}^N \to \mathbb{C}^N$ that describe the coupling between the layers i and j. We assume that we have fixed an orthonormal basis for the internal Hilbert space so that we can treat the blocks $H_{i,j}(\mathbf{k})$ as $N \times N$ matrices. From now on we will omit the dependence on $\mathbf{k}$ to make the notation clearer. We consider the case where there are only interactions inside each layer and between adjacent layers, and that these interactions are identical for each layer, so that we can write

$$H_{i,i} = V, \quad H_{i,i-1} = A, \quad H_{i,i+1} = A^\dagger, \quad \text{and } H_{i,j} = 0 \text{ else.} \tag{22}$$

We denote the vector describing the internal structure of the ith layer by $\psi_i \in \mathcal{H}_i$, and write $\psi = (\psi_1, \psi_2, \dots) \in \ell^2(\mathbb{N}) \otimes \mathbb{C}^N$. In what follows, I denotes the identity operator, where the space on which it acts follows clearly from the context. The Schrödinger equation describing the system is

$$(H - \varepsilon I)\psi = 0,$$

where ε denotes the energy of the state ψ. With equation (22), this takes the form of a second-order difference equation

$$-A\psi_{i-1} + \big(\varepsilon I - V\big)\psi_i - A^\dagger \psi_{i+1} = 0 \quad \text{for } i > 1, \tag{23}$$
$$\big(\varepsilon I - V\big)\psi_i - A^\dagger \psi_{i+1} = 0 \quad \text{for } i = 1, \tag{24}$$

where the latter equation can be considered as a Dirichlet boundary conditions by taking equation (23) and formally setting $\psi_0 = 0$. We now consider a Green's function matrix $G(\varepsilon, \mathbf{k})$ that obeys

$$(\varepsilon I - H)G = I.$$

We remark that when dealing with Green's functions in physics, it is customary to replace the energy ε by a complex energy $\varepsilon^+ = \varepsilon + i\eta$, where η is a small positive number. In this case one calls $G(\varepsilon^+, \mathbf{k})$ the retarded Green's function. Analytically, the limit $\eta \to 0$ is usually taken a later stage, and numerically, we keep $\eta > 0$ to allow for numerical computations. In this context η is referred to as the broadening parameter, because increasing η has the effect of broadening the peaks near the singularities in the Green's function.

As we did with the Hamiltonian, we subdivide the Green's function G into blocks $G_{i,j}$. These blocks then obey the relation

$$-AG_{i-1,j} + \big(\varepsilon I - V\big)G_{i,j} - A^\dagger G_{i+1,j} = \delta_{i,j}I, \quad i > 1.$$

where $\delta_{i,j}$ is the Kronecker delta. The diagonal blocks $G_{i,i}$ encode the the density of states of the ith layer. In particular, the density of states at the boundary layer $i = 1$ can be extracted from the block $G_{1,1}$. Therefore, this block is called the surface Green's function, denoted by

$$g(\varepsilon, \mathbf{k}) = G_{1,1}(\varepsilon, \mathbf{k}).$$

We now describe how an expression for the Green's function blocks $G_{i,i}(\varepsilon, \mathbf{k})$ can be obtained using the solutions to the Schrödinger equation. Because of the translational symmetry, we use the Bloch wave solution

$$\psi_{i+1} = \lambda \psi_i$$

for $i > 1$ as an ansatz, where λ is any complex number. Inserting this ansatz into equation (23) leads to the quadratic eigenvalue equation

$$\left[-A + \lambda(\varepsilon I - V) - \lambda^2 A^\dagger \right] u = 0,$$

with $u \in \mathbb{C}^N$. This equation can be solved by rewriting it to an equivalent generalized eigenvalue problem given by

$$\left(\begin{bmatrix} I & 0 \\ \varepsilon I - V & -A \end{bmatrix} - \lambda \begin{bmatrix} 0 & I \\ A^\dagger & 0 \end{bmatrix} \right) \begin{bmatrix} \lambda u \\ u \end{bmatrix} = 0.$$

In the at most $2M$ solutions $\{(\lambda_n, u_n)\}$ to this equation, we distinguish left-travelling waves and right-travelling waves. A first distinction can be made by considering evanescent waves that decay either to the left or to the right. Left-decaying waves are characterized by $|\lambda_n| > 1$, whereas right-decaying waves have $|\lambda_n| < 1$. This usually suffices, but if there are remaining solutions that satisfy $|\lambda_n| = 1$, one can compute the group velocity of the wave given by

$$v_n \sim -\text{Im}\left(\lambda_n u_n^\dagger A^\dagger u_n \right),$$

up to a positive constant factor. Left-travelling waves satisfy $v_n < 0$ and right-travelling waves satisfy $v_n > 0$. For the purpose of finding surface states, we are only interested in right-travelling waves, which we denote by $(\lambda_{n,R}, u_{n,R})$. Similarly, left-travelling waves are denoted by $(\lambda_{n,L}, u_{n,L})$ and we denote the number of left- and right-travelling waves by N_L and N_R respectively.

We now collect the left- and right-travelling waves into two matrices U_L and U_R given by

$$U_L = (u_{1,L}, \ldots, u_{N_L,L}), \quad U_R = (u_{1,R}, \ldots, u_{N_R,R}),$$

which are used to construct the so-called Bloch matrices,

$$F_L = U_L \Lambda_L U_L^+, \quad F_R = U_R \Lambda_R U_R^+,$$

where $\Lambda_{L/R} = \text{diag}(\lambda_{1,L/R}, \ldots, \lambda_{N_{L/R},L/R})$ and M^+ denotes the pseudoinverse of a matrix M. It can be shown [23] that the surface Green's function is then the $N \times N$ matrix-valued function given by

$$g(\varepsilon, \mathbf{k}) = \left[A F_R^+ \right]^{-1},$$

where we recall that the right-hand side depends on $\mathbf{k}$ and ε. In the case that the matrix in brackets is singular, a pseudoinversion is understood. The Green's function of the ith layer in the bulk can be expressed as

$$G_{i,i} = \left[A^\dagger F_L - A^\dagger F_R \right]^{-1}.$$

The density of states at the ith layer, denoted by n_i, is related to the Green's function blocks by

$$n_i(\varepsilon, \mathbf{k}) = -\frac{1}{\pi} \text{Im}\left(\text{Tr } G_{i,i}(\varepsilon, \mathbf{k}) \right).$$

In particular, the surface density of states $n_S = n_1$ is given by

$$n_S(\varepsilon, \mathbf{k}) = -\frac{1}{\pi} \text{Im}\left(\text{Tr } g(\varepsilon, \mathbf{k}) \right).$$

The density of states in the bulk layers encodes the properties of the electrons inside the bulk, but the surface density of state reveals states that emerge only at the surface.

5.2 Edge States in the BHZ Model

We now apply the theory of surface Green's functions reviewed in the previous subsection to the two-dimensional BHZ model that has been introduced in section 3. In this two-dimensional context, we refer to $g(\varepsilon, k_y)$ as the edge Green's function. Recall that the Bloch Hamiltonian of the BHZ model is given by

$$H_{\mathrm{BHZ}}(k_x, k_y) = s_0 \otimes \left[(u + \cos k_x + \cos k_y)\sigma_z + \sin k_y \sigma_y\right] + s_z \otimes \sin k_x \sigma_x + s_x \otimes C,$$

where we set the coupling operator to $C = 0.3\sigma_y$ as before. We consider a semi-infinite square lattice with an edge parallel to the y-direction, as in figure 10. The first step that we have to take is to reconstruct the lattice Hamiltonian along the x-direction, which is no longer translationally invariant. This can be done by starting with the Bloch Hamiltonian for k_x and k_y describing a fully translationally invariant lattice and using equation (6) restricted to the x-direction, so that the partial Bloch Hamiltonian for k_y is given by

$$H(k_y) = \sum_{k_x} |k_x\rangle \langle k_x| \otimes H(k_x, k_y),$$

where k_x runs over all plane waves along the x-direction. In appendix C we show that the expression for the Bloch Hamiltonian $H(k_y)$ becomes

$$
\begin{aligned}
H(k_y) = & \frac{1}{2}\sum_{i_x} \left[|i_x + 1\rangle \langle i_x| \otimes s_0 \otimes \sigma_z + |i_x\rangle \langle i_x + 1| \otimes s_0 \otimes \sigma_z \right] \\
& + \frac{i}{2}\sum_{i_x} \left[|i_x + 1\rangle \langle i_x| \otimes s_z \otimes \sigma_x - |i_x\rangle \langle i_x + 1| \otimes s_z \otimes \sigma_x \right] \\
& + \sum_{i_x} |i_x\rangle \langle i_x| \otimes \left[s_x \otimes C + s_0 \otimes \left[(u + \cos k_y)\sigma_z + \sin k_y \sigma_y\right] \right],
\end{aligned}
$$

where i_x temporarily denotes the layer index to distinguish it from the imaginary unit. Comparing to equation (22), we find that the operators V and A acting between the layers are given by

$$V = s_x \otimes C + s_0 \otimes \left[(u + \cos k_y)\sigma_z + \sin k_y \sigma_y\right], \quad A = \frac{1}{2}s_0 \otimes \sigma_z + \frac{i}{2}s_z \otimes \sigma_x.$$

The scheme to compute the density of states described in the previous subsection can now be carried out numerically. The numerical computations have been done using MATLAB.

We compute the edge density of states for $u = -2.8$ and $u = -1.2$. The results of these computations are shown in figure 11. (a) and (b) correspond to the case $u = -2.8$, for which we have seen that the $\mathbb{Z}_2$-index is trivial. Indeed, no gapless edge states are observed, and the edge density of states resembles the bulk density of states. On the other hand, for (c) and (d) we have set $u = -1.2$, which gives a non-trivial $\mathbb{Z}_2$-index as we have seen in the previous section. The edge density of states in (d) indeed shows the existence of gapless edge states. The two crossing branches near $k_y = 0$ are the signature of two-dimensional time-reversal symmetric topological insulators. The fact that these states are truly localized at the edge of the system can be deduced by comparing with (c), which shows that the bulk layers have a band gap. The branches cross in a Dirac point, where the dispersion relation $\varepsilon(k_y)$ is linear.
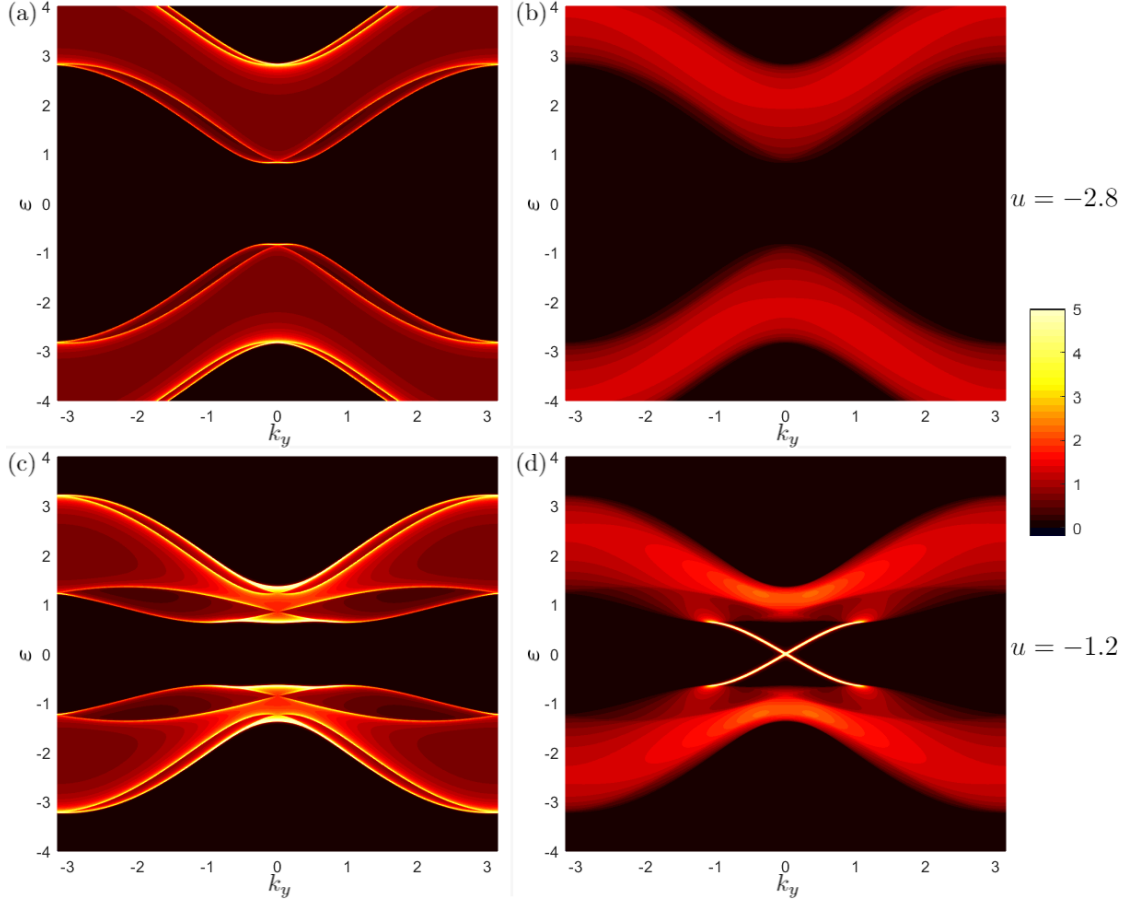
FIGURE 11: Layer-resolved density of states (DOS) $n(\varepsilon, k_y)$ as a function of energy ε and wavenumber $k_y \in [-\pi, \pi]$ of the BHZ model. Warmer colors indicate a higher density. (a) Bulk layer DOS for the trivial case $u = -2.8$. (b) Edge layer DOS for the trivial case $u = -2.8$. (c) Bulk layer DOS for the non-trivial case $u = -1.2$. (d) Edge layer DOS for the non-trivial case $u = -1.2$, showing gapless edge states. The broadening parameter is set to $\eta = 0.005$ in all computations. To visualize the DOS, it is capped at 5. Note that ε and k_y are dimensionless in the BHZ model.

5.3 Surface States in Bi₂Se₃

Now that we have studied the existence of topologically protected gapless edge states in a simple model, we move back to our three-dimensional model of Bi₂Se₃ and analyze the electronic properties that emerge at the surface. We consider a semi-infinite lattice subdivided into quintuple layers, as in figure 7. In the coordinate system chosen in that figure, the quintuple layers are stacked in the z-direction, so that the surface is parallel to the xy-plane. This is not only the most simple surface to study, but also the most natural one, as the planes separating the quintuple layers are cleavage planes[11] of Bi₂Se₃. It is called the [111] surface, which refers to the chosen primitive lattice vectors. With regards to the labelling of the atomic sites in the unit cell that we introduced, each type of atomic site occurs precisely once in each quintuple layer, ordered as $(2, 5, 1, 4, 3)$. We illustrate the setup in figure 12.

---

[11]The bonds between the quintuple layers are mainly of van der Waals type, and they are relatively weak so that these bonds are easily broken in Bi₂Se₃ crystals.
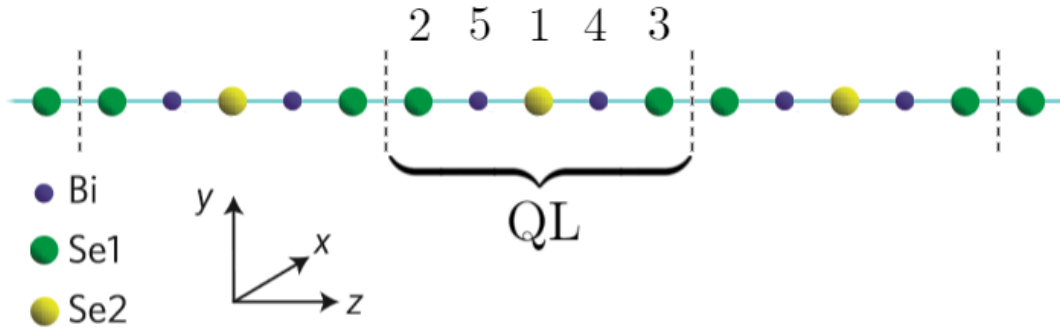
FIGURE 12: Subdivision of the crystal structure of Bi_2Se_3 into quintuple layers(QLs) stacked along the z-direction. The dashed lines indicate the bounds of the layers. Each atom in this figure represents an infinite plane of atoms in a triangular lattice, stacked as in figure 7 in an ABCABC stacking sequence. Note that the blue lines do not represent the bonds between atoms. Adapted from [15].

The matrix blocks V and A appearing in equation (22) can be extracted from the Bloch Hamiltonian $H(\mathbf{k})$ considered in section 4 by subdividing the interactions of the atoms into interactions taking place within the quintuple layers and interactions coupling two adjacent quintuple layers. From figure 12 we see that the nearest neighbour and second nearest neighbour interactions between adjacent layers correspond to $3 \leftrightarrow 2$, $3 \leftrightarrow 5$ and $4 \leftrightarrow 2$. Hence, the interaction blocks A and $A^\dagger$ are the blocks of $H(\mathbf{k})$ corresponding to these interactions, and the remaining blocks belong to V. The surface density of states can now be computed as a function of energy ε and the wavevector $\mathbf{k} = (k_x, k_y)$ parallel to the surface using the method of surface Green's functions. The results of the computations are shown in figure 13.
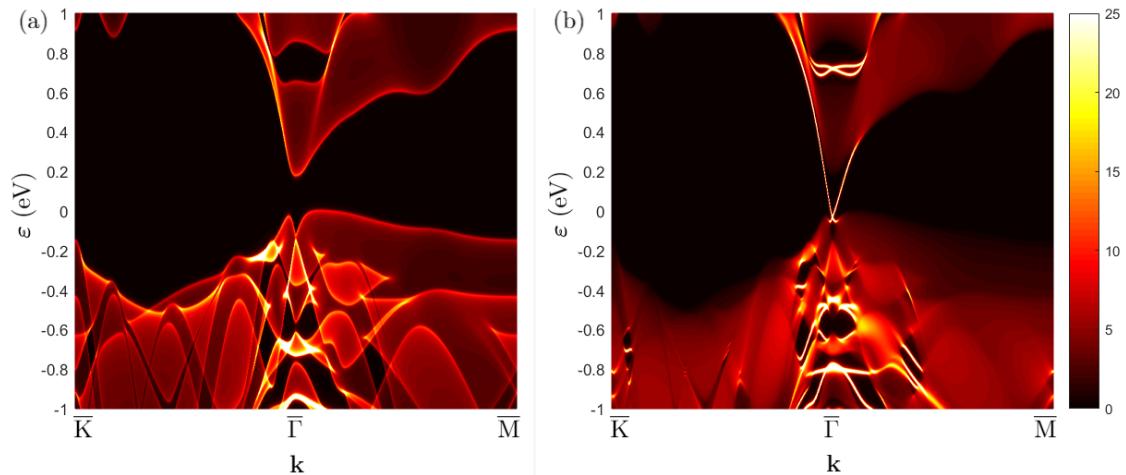
FIGURE 13: Layer-resolved density of states (DOS) $n(\varepsilon, \mathbf{k})$ of Bi_2Se_3 as a function of energy ε (eV) and wavenumber $\mathbf{k}$ on the high-symmetry path $\overline{K}$-$\overline{\Gamma}$-$\overline{M}$. Warmer colors indicate a higher density. (a) Bulk layer DOS, showing a band gap at the Fermi level. (b) Surface layer DOS on the [111] surface, with a single surface Dirac cone emerging from the top of the valence band. The broadening parameter is set to $\eta = 0.005$ in all computations. To visualize the DOS, it is capped at 25.

38

Figure 13 confirms that there are topologically protected surface states on the surface of Bi_2Se_3, and a Dirac cone can be observed on the surface, traversing the bulk band gap. In figure 14 we zoom in into the Dirac cone and give a three-dimensional view of the surface density of states near the $\overline{\Gamma}$ point.

FIGURE 14: Dirac cone on the surface of Bi_2Se_3 at the $\overline{\Gamma}$-point, encoded in the surface density of states $n(\varepsilon, \mathbf{k})$. (a) Cross section of the Dirac cone, with $\mathbf{k}$ tracing a segment of the high-symmetry path $\overline{M}$-$\overline{\Gamma}$-$\overline{M}$ centered at the $\overline{\Gamma}$-point in units of $\mathring{A}^{-1}$. (b) Three-dimensional view of the Dirac cone. The broadening parameter is set to $\eta = 0.0006$ and the DOS is capped at 1000.
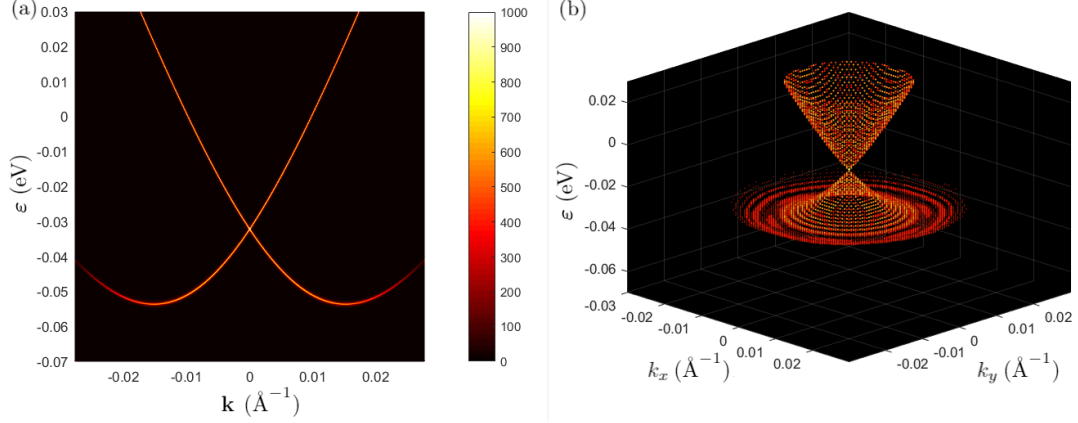
This single Dirac cone on the surface is the signature of a three-dimensional topological insulator. The linear dispersion relation near the Dirac point accounts for the fact that the surface electrons can behave like massless particles obeying the Dirac equation. The Fermi velocity v_F of the states in the Dirac cone, determined by the slope of the Dirac cone using

$$v_F = \frac{1}{\hbar} \frac{\partial \varepsilon(k)}{\partial k}$$

evaluated at the Fermi level, is found to be $v_F \approx 1.4 \times 10^5$ m s^{-1}, which is in the same order of magnitude as found in [15]. The spin-momentum locking is not visible in figure 14, and more work is required to extract the spin expectation values of the surface states from the surface Green's function.

The results agree qualitatively with experiments, which also show that there is a single Dirac cone on the surface of Bi_2Se_3 [14]. Using angle-resolved photoemission spectroscopy (ARPES), the momentum-space electronic structure on the surface of a sample can be measured directly. Figure 15 shows the results of ARPES measurements on the [111] surface of Bi_2Se_3, which is the same surface as the one that we considered.
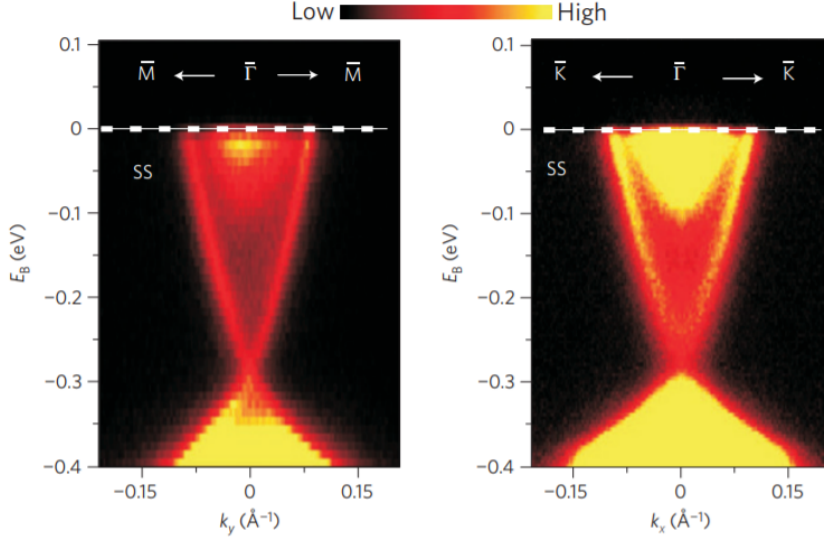
FIGURE 15: ARPES measurements of the [111] surface of a Bi_2Se_3 sample, showing the electronic structure near the $\overline{\Gamma}$ point as a function of binding energy E_B (eV) and momentum k_y (left) and k_x (right) (Å^{-1}). From [14].

To conclude this section, this shows that a simple tight-binding model is able to capture the topological behaviour on the surface of a material in a realistic way. It shows that the computation of the $\mathbb{Z}_2$-index, whose formulation is relatively abstract, indeed allows one to predict potentially interesting surface phenomena that can be observed directly by experiments. The principle behind this is the bulk-boundary correspondence, and in the next section we discuss some of the mathematical principles behind this correspondence.

6 The Bulk-Edge Correspondence

In the preceding sections we have seen that the topological $\mathbb{Z}_2$-index of the bulk Hamiltonian determines the existence of topologically protected surface and edge states, by computing $\mathbb{Z}_2$-indices of different tight-binding models and calculating the density of states at the boundary of semi-infinite lattices. This is a manifestation of the bulk-boundary correspondence. In this section we review a mathematical version of the bulk-boundary correspondence for two-dimensional time-reversal symmetric topological insulators, and we outline its proof which is formulated by Graf and Porta [19], who prove the result in a similar setting as the one we considered in section 5. In this context the theorem is referred to as the bulk-edge correspondence. The full proof is quite lengthy and technical, so only the main steps are given. The aim of this section is to give the reader an overview of the mathematical ideas behind the bulk-edge correspondence without making the discussion too formal. We refer the reader who is interested in the detailed proof to [19].

6.1 Setting of the Theorem

We first introduce the setting in which we will formulate the bulk-edge correspondence, which is similar to that of the previous sections. We consider a lattice $\mathbb{Z} \times \mathbb{Z}$, which we regard as the bulk system, and a semi-infinite lattice $\mathbb{N} \times \mathbb{Z}$, which we view as the edge version of the latter system. This may seem like a restricted setting, but in principle any two-dimensional crystal structure can be encoded in this lattice, if we view the lattice sites as labels rather than locations of atoms. We consider a tight-binding Hamiltonian H with N internal degrees of freedom on the $\mathbb{Z} \times \mathbb{Z}$ lattice, assuming translational invariance in the second direction so that we can use Bloch's theorem to obtain a family of Bloch Hamiltonians $H(k)$ parametrized by k in the Brillouin zone $\mathfrak{B} = S^1 = \mathbb{R}/2\pi\mathbb{Z}$. This Hamiltonian, which we will call the bulk Hamiltonian, is assumed to act on states $\psi = (\psi_n)_{n \in \mathbb{Z}} \in \ell^2(\mathbb{Z}) \otimes \mathbb{C}^N$ as

$$(H(k)\psi)_n = A(k)\psi_{n-1} + V(k)\psi_n + A(k)^\dagger \psi_{n+1}, \tag{25}$$

where $A(k)$ and $V(k)$ are $N \times N$ complex matrices, with each $A(k)$ invertible and each $V(k)$ Hermitian. To this Hamiltonian we associate an edge Hamiltonian $H^\sharp(k)$, which acts on $\psi^\sharp = (\psi_n^\sharp)_{n \in \mathbb{N}} \in \ell^2(\mathbb{N}) \otimes \mathbb{C}^N$ as

$$(H^\sharp(k)\psi^\sharp)_n = A(k)\psi_{n-1}^\sharp + V(k)\psi_n^\sharp + A(k)^\dagger \psi_{n+1}^\sharp, \tag{26}$$

where the above equation at $n = 1$ is to be read with $\psi_0^\sharp = 0$, to be interpreted as a boundary condition. For simplicity we consider the bulk and edge Hamiltonian to be identical except for the boundary condition on $H^\sharp(k)$, but in [19] the edge Hamiltonian is allowed to be different within a finite distance from the edge to incorporate possible edge effects. Since we are interested in insulators, the bulk Hamiltonian should have a band gap at the Fermi level, so we assume that

$$\varepsilon_{\mathrm{F}} \notin \sigma(H(k)), \quad \text{for all } k \in \mathfrak{B}.$$

Furthermore, we consider the case where $H(k)$ and $H^\sharp(k)$ have time-reversal symmetry, represented by an operator Θ, which is assumed to have the following properties:

(i) Θ is anti-linear and $\Theta^2 = -1$;

(ii) $\Theta^\dagger \Theta = 1$;

(iii) For all $k \in \mathfrak{B}$,

$$H(-k) = \Theta H(k) \Theta^{-1}. \tag{27}$$

Within this setup we introduce a bulk index $\mathcal{I} \in \{\pm 1\}$ and an edge index $\mathcal{I}^{\sharp} \in \{\pm 1\}$. The bulk index is equivalent to the $\mathbb{Z}_2$-index and is defined in terms of $H(k)$, and $\mathcal{I}^{\sharp}$ will be the parity of the number of pairs of edge states. The bulk-edge correspondence is then simply expressed as $\mathcal{I} = \mathcal{I}^{\sharp}$, which is the main result whose proof we will outline in this section. The next subsection is devoted to the construction of $\mathcal{I}$.

6.2 Three Auxiliary Indices

Here we construct the bulk index $\mathcal{I}$ from a given bulk Hamiltonian $H(k)$. Although the Schrödinger equation

$$(H(k) - \varepsilon)\psi = 0$$

is an equation in the Hilbert space of the system, we may also regard it as a second-order difference equation in $n \in \mathbb{Z}$, similar to the difference equation considered in section 5, and consider solutions that live outside $\ell^2(\mathbb{Z}) \otimes \mathbb{C}^N$. It can be shown [19] that if $\varepsilon \in \mathbb{C}$ is not an eigenvalue of the Hamiltonian $H(k)$, then this difference equation still has $2N$ linearly independent solutions in $\mathbb{Z} \times \mathbb{C}^N$, of which N decay as $n \to \infty$. Hence, let $E_{\varepsilon,k}$ be the N-dimensional vector space of solutions that decay as $n \to \infty$ for a given wavenumber k energy ε outside the spectrum of $H(k)$. The time-reversal symmetry of the Hamiltonian implies that

$$\big(H(-k) - \overline{\varepsilon}\big)\Theta\psi = \Theta\big(H(k) - \varepsilon\big)\psi.$$

Hence, if ψ solves the Schrödinger equation at (ε, k), then the time-reversed state $\Theta\psi$ solves the Schrödinger equation at $(\overline{\varepsilon}, -k)$. In other words, the solution spaces are related by $E_{\overline{\varepsilon}, -k} = \Theta E_{\varepsilon,k}$. In the following we will consider complex-valued energies, so we change the notation from ε to z. Now, let γ be a contour in $\mathbb{C}$ with $\gamma = \overline{\gamma}$ that encloses the the part of the bulk spectrum $\sigma(H(k))$ below the Fermi energy ε_{F} for every $k \in S^1$. Such a contour γ exists because of the assumption that the spectrum is gapped at the Fermi energy. Let $\mathbb{T} = \gamma \times S^1$, which is a torus. Figure 16 shows an illustration of this setup. Because of this choice of γ, we have an involution[12] τ on the torus $\mathbb{T}$ given by $(z, k) \mapsto (\overline{z}, -k)$.

---

[12]An involution is a map which is equal to its own inverse. In other words, if τ is an involution, then $\tau^2 = 1$.
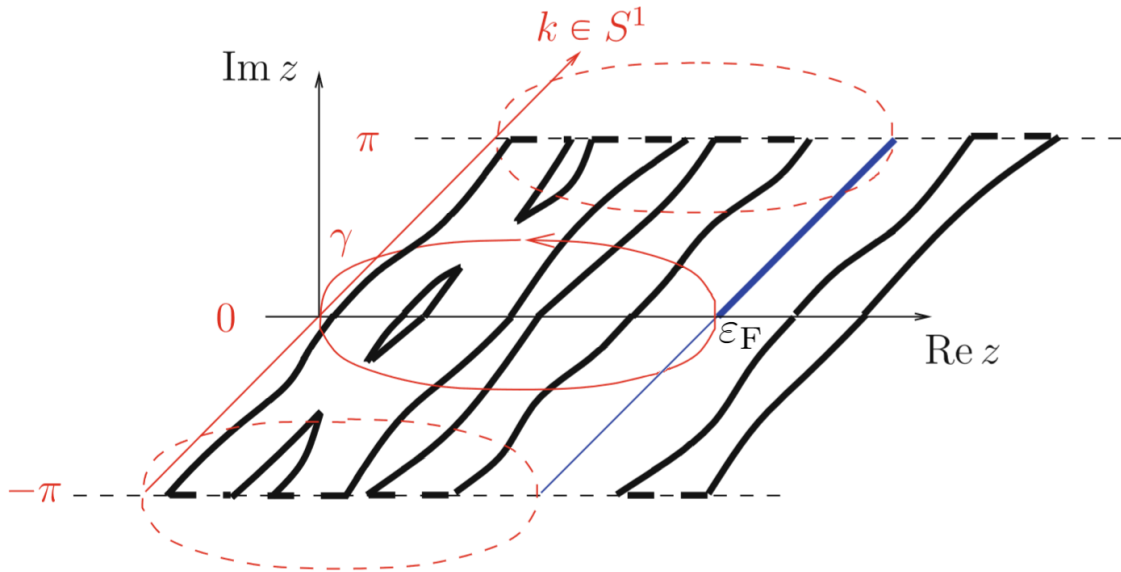
FIGURE 16: Spectrum $\sigma(H(k)) \subset \mathbb{C}$ for k on the Brillouin zone circle S^1, indicated by the thick black lines. The circle is cut open at $\pi \equiv -\pi$, so the dashed red loops are to be identified. Because $H(k)$ is Hermitian, the spectrum lies on the real line. The Fermi energy ε_{F} is represented by the blue line. The contour γ is reflection symmetric and encircles the part of the bulk spectrum with energies below ε_{F}. Adapted from Graf and Porta [19].

We now collect all the solution spaces $E_{z,k}$ together and set

$$E = \coprod_{(z,k) \in \mathbb{T}} E_{z,k},$$

where $\coprod$ is the disjoint union[13]. Together with the canonical projection map $E \to \mathbb{T}$ that sends a solution in $E_{z,k}$ to the basepoint $(z, k) \in \mathbb{T}$, this space becomes a vector bundle with fibers $E_{z,k}$. From this bundle we can construct the bulk index $\mathcal{I}$, which we will do step by step by constructing three auxiliary indices.

Index of Endpoint Degenerate Families

First, we consider an index defined for families of complex numbers that move continuously over the circle S^1 as a function as a function of some parameter with values in $[0, \pi]$. These continuous families are of the form $\{Z(\phi) : \phi \in [0, \pi]\}$, where $Z(\phi) = (z_1(\phi), \ldots, z_N(\phi))$, with each z_i a continuous map $[0, \pi] \to S^1$. We note that the motion of Wannier charge centers over the circle is an example of such a continuous family. Generically[14], we can choose a point p on the circle S^1 so that the $(z_1(\phi), \ldots, z_N(\phi))$ cross p only finitely many times, i.e. $p \in Z(\phi)$ for finitely many ϕ, such that each of the crossings is *simple* in the sense that $p = z_j(\phi)$ for at most one j and $z_j'(\phi) \neq 0$. Suppose that we have such a family Z which is *endpoint degenerate*, meaning that each point belonging to the endpoints $Z(a)$ or $Z(b)$ occur an even number of times. We then define an index $\mathcal{I}_{\mathrm{I}}$ for such Z by

$$\mathcal{I}_{\mathrm{I}}(Z) := (-1)^n,$$

---

[13]Intuitively, the disjoint union is a way of taking the union of sets without letting them overlap if they have common elements.

[14]The term *generic* has a precise meaning in this context. A property of some object is called generic if it satisfied for an open dense subset of the space to which this object belongs. We will not discuss the proof at this level of formality, so the term can be read loosely.

where n is the number of times that the lines in $Z(\phi)$ cross p. The condition that $Z(\phi)$ is endpoint degenerate makes this index well-defined[15]. An example of an index degenerate family is shown in figure 17.
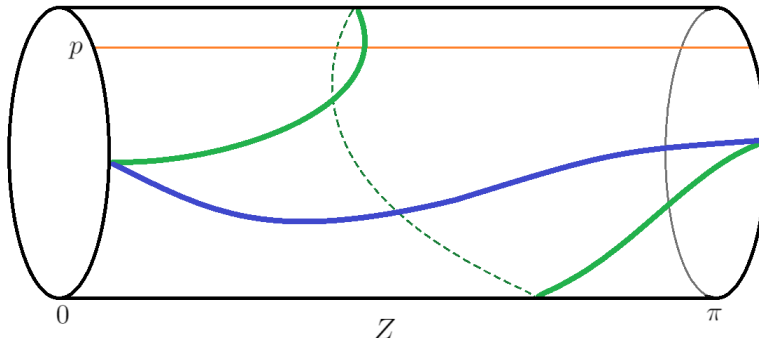

FIGURE 17: Continuous endpoint degenerate family $Z(\phi)$ on the unit circle S^1 with ϕ in $[0, \pi]$ (blue and green). The figure takes the shape of a cylinder because it represents a circle at each $\phi \in [0, \pi]$. Hence, the point p on the circle becomes a line (orange). It is crossed once, so that $n = 1$ and $\mathcal{I}_{\mathrm{I}}(Z) = -1$.

This index is directly related to the flow of the Wannier charge centers over the unit circle as a function of time, and $\mathcal{I}_I$ captures the same information as the $\mathbb{Z}_2$-index.

Index of Kramers Matrix Families

We now introduce a second auxiliary index based on the previous one. Let J be the anti-symmetric $N \times N$ block-diagonal matrix whose blocks are given by

$$\begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

and let $K : \mathbb{C}^N \to \mathbb{C}^N$ be complex conjugation. We call $\Theta_0 := JK$ the *standard time-reversal* on $\mathbb{C}^N$.

Suppose that we have a continuous family of invertible matrices $T(\phi) \in \mathrm{GL}(N, \mathbb{C})$ parametrized by $\phi \in [0, \pi]$. Then T is said to have the *Kramers property* if

$$\Theta_0 T(0) = T(0)^{-1}\Theta_0 \quad \text{and} \quad \Theta_0 T(\pi) = T(\pi)^{-1}\Theta_0. \tag{28}$$

This property is not to be confused with time-reversal symmetry and Kramers degeneracy, although it has similar properties. Analogous to Kramers degeneracy, it can be shown that the eigenvalues of $T(0)$ and $T(\pi)$ come in pairs $\lambda, \overline{\lambda}^{-1}$. In the case that $\lambda = \overline{\lambda}^{-1}$, these eigenvalues are doubly degenerate. This means in particular that their phases $z = \lambda/|\lambda|$ are always evenly degenerate. Hence, if we let $\lambda_i(\phi)$ denote the eigenvalues of $T(\phi)$, which vary continuously with ϕ, then setting

$$z_i(\phi) = \frac{\lambda_i(\phi)}{|\lambda_i(\phi)|}$$

---

[15]We remark that Graf and Porta construct this index in a more robust way using a winding number, which can also be defined for endpoint degenerate families Z for which such a point p does not exist. This construction is quite technical and does not significantly contribute to the understanding of the proof.

gives rise to a continuous endpoint degenerate family $Z(\phi) = (z_1(\phi), \ldots, z_N(\phi))$ of points on the unit circle S^1. We can therefore define an index $\mathcal{I}_{\mathrm{II}}$ for families of matrices T with the Kramers property by setting

$$\mathcal{I}_{\mathrm{II}}(T) := \mathcal{I}_{\mathrm{I}}(Z),$$

with Z as above. To give a link with earlier sections we remark that the Wannier charge centers were also obtained in terms of eigenvalues of a family of matrices, namely the Wilson loops Λ. We will soon see that this analogy can be pushed further.

Index of Time-Reversal Symmetric Bundles

We now define a topological index for a specific type of vector bundle that we define here. Let $\mathbb{T} = S^1 \times S^1$ be the torus, whose points we denote by $\varphi = (\varphi_1, \varphi_2)$. Let $\tau : \mathbb{T} \to \mathbb{T}$ be the involution $\varphi \mapsto -\varphi$. A vector bundle $\pi : E \to \mathbb{T}$ with fibers $E_\varphi = \pi^{-1}(\varphi)$ over the base space $\mathbb{T}$ is called *time-reversal symmetric* if there is a map $\Theta : E \to E$ satisfying $\Theta^2 = -1$ which maps the fiber E_φ anti-linearly onto the fiber $E_{\tau\varphi}$. In other words, the fibers of E come in pairs related by time-reversal through the involution τ.

Let $F(E)$ be the frame bundle[16] of E, whose elements we denote by $v = (v_1, \ldots, v_M) \in F(E)_\varphi$ for a given φ. The frame bundle $F(E)$ comes naturally with a right action by the group $\mathrm{GL}(N, \mathbb{C})$ of invertible matrices, given by $M : F(E)_\varphi \to F(E)_\varphi$, $v \mapsto vM$ on the fibers. Any two frames are related by this action via a unique invertible matrix M, namely the corresponding change-of-basis matrix. If E is time-reversal invariant, $F(E)$ also admits a time-reversal map Θ acting as

$$\Theta : F(E)_\varphi \to F(E)_{\tau\varphi}$$
$$v \mapsto \Theta v,$$

where $(\Theta v)_i = \Theta v_i$.

For the definition of the third auxiliary index, we now cut the torus along the circle $\{\pi\} \times S^1$ to obtain a cut torus $\widetilde{\mathbb{T}} = [-\pi, \pi] \times S^1$. Any vector bundle E with base space $\mathbb{T}$ naturally gives rise to a vector bundle $\widetilde{E}$ over $\widetilde{\mathbb{T}}$. Using a construction based on homotopy arguments, it can be shown that the frame bundle $F(\widetilde{E})$ admits a smooth section[17] $v : \widetilde{\mathbb{T}} \to F(\widetilde{E})$ that satisfies the time-reversal condition

$$v(\tau\varphi) = \Theta v(\varphi) J.$$

Given such a section v, we set $v_\pm(\varphi_2) = v(\pm\pi, \varphi_2)$, referred to as the boundary values. Because the fibers $E_{-\pi,\varphi_2}$ and E_{π,φ_2} are identical, there exists a unique invertible matrix $T(\varphi_2)$ for each φ_2 that relates the frames $v_+(\varphi_2)$ and $v_-(\varphi_2)$. That is, $T(\varphi_2)$ satisfies

$$v_+(\varphi_2) = v_-(\varphi_2) T(\varphi_2). \tag{29}$$

We call $T(\varphi_2)$ the transition matrix. It is analogous to the Wilson loop, as it relates two frames at a point on the circle, where one of the frames has been parallel transported

---

[16]A frame of a vector space is an ordered basis, and the frame bundle of a given vector bundle is the bundle whose fibers $F(E)_\varphi$ are the spaces containing all frames of E_φ.

[17]A section σ of a bundle is a map from the base space into the bundle, such that the points on the base space are mapped into the fibers directly above it. In other words, $\pi \circ \sigma$ is the identity map. The existence of a smooth section of the frame bundle means that a continuously varying basis exists for each fiber, analogous to a smooth gauge of eigenstates.

around a circle. One can show, relatively straightforwardly, that these transition matrices $T(\varphi_2)$ satisfy

$$\Theta_0 T(\varphi_2) = T(-\varphi_2)^{-1} \Theta_0. \tag{30}$$

In other words, the family $T(\varphi_2)$ has the Kramers property on $\varphi_2 \in [0, \pi]$. This mean that we can assign the index $\mathcal{I}_{\text{II}}$ to $T(\varphi_2)$. A key intermediate result is that the index $\mathcal{I}_{\text{II}}(T)$ is actually independent of the chosen time-reversal symmetric section v, which can be shown using homotopy arguments. Because of this independence, the index $\mathcal{I}_{\text{II}}(T)$ is intrinsic to the bundle E, which allows us to define a third index $\mathcal{I}_{\text{III}}$ for a time-reversal symmetric bundle as

$$\mathcal{I}_{\text{III}}(E) := \mathcal{I}_{\text{II}}(T).$$

This concludes the preliminaries of the bulk-edge correspondence.

6.3 Bulk Index $\mathcal{I}$ and Edge Index $\mathcal{I}^\sharp$

With the three auxiliary indices defined we return to the setting of the theorem, where we had a bundle E over a torus $\mathbb{T} = \gamma \times S^1$ whose fibers $E_{z,k}$ consisted of certain solutions to the Schrödinger equation. Because of the property $E_{\bar{z},-k} = \Theta E_{z,k}$, the bundle E is in fact a time-reversal symmetric bundle in the sense of the previous subsection. Therefore, it has a well defined index $\mathcal{I}_{\text{III}}$. This finally leads to the definition of the bulk index, which we define to be

$$\mathcal{I} := \mathcal{I}_{\text{III}}(E).$$

As we remarked earlier, this index is equivalent to the $\mathbb{Z}_2$-index, as shown in [19]. For the edge index, consider a small interval I containing the Fermi energy that lies in the band gap, i.e. $\sigma(H(k)) \cap I = \emptyset$ for all k. The part of the edge spectrum $\sigma(H^\sharp(k))$ that lies in I consists of continuous and differentiable eigenvalue branches $\varepsilon_i(k)$, and it may occur that these eigenvalue branches cross the Fermi level, so $\varepsilon_i(k_*) = \varepsilon_{\text{F}}$ for some $k_* \in S^1$. We may assume that these eigenvalue crossings are simple, meaning that $\varepsilon_i'(k_*) \neq 0$ for the crossing points k_*. If n is the number of crossing points k_* in the interval $[0, \pi]$, with the crossings at $k_* = 0$ and $k_* = \pi$ counted half to account for Kramers degeneracy, we define the edge index to be

$$\mathcal{I}^\sharp = (-1)^n.$$

In analogy with our discussion in section 3, $\mathcal{I}^\sharp$ is the parity of the number of Kramers pairs of edge states, with $\mathcal{I}^\sharp = -1$ indicating the presence of topologically protected edge states. With the bulk and edge index defined, we can finally formulate the main theorem.

Theorem [Bulk-Edge Correspondence] With the bulk and edge Hamiltonian defined as in the beginning of this section, assumed to be time-reversal symmetric and spectrally gapped at the Fermi level, we have

$$\mathcal{I} = \mathcal{I}^\sharp.$$

6.4 $\mathcal{I} = \mathcal{I}^\sharp$ - Outline of the Proof

As we remarked, the full proof of the bulk-edge correspondence is too lengthy to discuss here in detail. Hence, we will give the main steps.

1. Bulk frames and edge solutions

Because the Hamiltonians $H(k)$ and $H^\sharp(k)$ act identically except for the boundary condition $\psi_0^\sharp = 0$ for the edge Hamiltonian, a solution $\psi \in \mathbb{Z} \times \mathbb{C}^N$ to $(H(k) - z)\psi = 0$ also satisfies $(H^\sharp(k) - z)\psi = 0$ without the boundary condition. A frame $\Psi \in F(E)_{z,k}$ is an ordered set $(\psi^1, \ldots, \psi^N)$ of linearly independent solutions in $E_{z,k}$, which we may equivalently view as a sequence of $N \times N$ complex matrices $\Psi = (\Psi_n)$ indexed by $n \in \mathbb{Z}$ that satisfies the matrix equation

$$(H(k) - z)\Psi = 0$$

with the property that Ψ_n decays as $n \to \infty$. Hence, any edge solution $\psi^\sharp$ to

$$(H^\sharp(k) - z)\psi^\sharp = 0,$$

with or without boundary condition, can be expressed in terms of a frame Ψ as $\psi^\sharp = \Psi a$ for some complex vector a.

2. Edge states in terms of singularity of Ψ_0

For edge states, we are interested in eigenvalues of $H^\sharp$, which means that we look for edge solutions $\psi^\sharp \in \mathbb{Z} \times \mathbb{C}^N$ that also satisfy the boundary condition $\psi_0^\sharp = 0$. It can be shown relatively directly that this boundary condition is satisfied precisely when Ψ_0 is singular. Moreover, the only points $(z, k) \in \mathbb{T}$ for which this can be the case are those that have $z = \varepsilon_{\mathrm{F}}$. Using perturbation theory, it can be shown that the points k_* in $(\varepsilon_{\mathrm{F}}, k_*)$ for which the latter holds can be shown to be generically isolated in the Brillouin zone S^1. By generically we mean that this property holds for a dense set of Hamiltonians near the given one: if the given Hamiltonian does not have this property, then we can approximate our Hamiltonian arbitrarily closely, and prove the result for the set of approximating Hamiltonians. The last part of this step amounts to showing that the crossing points $(\varepsilon_{\mathrm{F}}, k_*)$ are given by simple eigenvalue branches $\varepsilon(k)$ of the edge Hamiltonian $H^\sharp(k)$.

3. Edge states encoded in a family of matrices $L(z, k)$

If Ψ is a local section of the frame bundle near a crossing point $(\varepsilon_{\mathrm{F}}, k_*)$, we can define a family of matrices given by

$$L(z, k) = -\Psi_1^\dagger(\bar{z}, k) A(k) \Psi_0(z, k).$$

Using the basic properties of frames Ψ and the difference equation that they satisfy, it can be shown without too much effort that this family of matrices satisfies $L(z, k) = L(\bar{z}, k)^\dagger$, so that it has real eigenvalues if z is real. From what has been shown in the first step, we see that $L(z, k)$ has zero as an eigenvalue precisely when (z, k) admits an edge state solution: $\Psi_1(z, k)$ and $A(k)$ are invertible for any k, and $\Psi_0(z, k)$ is singular if and only if $(z, k) = (\varepsilon_{\mathrm{F}}, k_*)$ is a crossing point of the edge spectrum with the Fermi line. Generically, $L(z, k)$ has a single eigenvalue branch $l(z, k)$ that vanishes precisely at the these crossing points, and the derivatives $\partial l / \partial z$ and $\partial l / \partial k$ can be shown to be real and non-zero.

4. Construction of a time-reversal invariant section

The next step is to construct a specific section of frames for any of the generic Hamiltonians for which the properties of the first step hold. We want this section to have certain properties for use in later steps. For (z, k) away from any crossing point, we can define a section Ψ through the condition

$$\Psi_0(z, k) = 1.$$

In other words, we choose the section Ψ such that the 0th matrix Ψ_0 in the sequence $\Psi = (\Psi_n)$ is the identity matrix. By the first step, this condition cannot be imposed at any of the crossing points (ε_F, k_*). Instead, an alternative local section $\widetilde{\Psi}$ can be defined inside a small disk $D \subset \mathbb{T}$ containing a crossing point (ε_F, k_*) in its interior, through the condition

$$\widetilde{\Psi}_1(z, k) = 1.$$

We recall that the bulk index $\mathcal{I}$ was defined in terms of a cut torus. Hence, we cut the torus open along the Fermi line $\{\varepsilon_F\} \times S^1$, obtaining the cut torus $\widetilde{\mathbb{T}}$. We now want to extend the section Ψ, which is not yet defined near the crossing points, to the cut torus $\widetilde{\mathbb{T}}$. We will extend Ψ based on the section $\widetilde{\Psi}$ defined near a crossing point (ε_F, k_*). Let $(\partial D)_\pm = \{z \in D : \pm \mathrm{Im}\, z > 0\}$ denote the upper and lower half boundary of the disk. On this boundary, both Ψ and $\widetilde{\Psi}$ are defined, and since they are frames, they must be related by

$$\widetilde{\Psi}(z, k) = \Psi(z, k) M_\pm(k) \quad \text{for } (z, k) \in (\partial D)_\pm,$$

where the $M_\pm(k)$ are invertible matrices parametrized by k in the interval I which is the intersection of the Fermi line with D. By defining

$$\Psi(z, k) = \widetilde{\Psi}(z, k) M_\pm(k)^{-1} \quad \text{for } (z, k) \in D_\pm,$$

we obtain a continuous section globally defined on $\widetilde{\mathbb{T}}$.

5. Making the problem local

To show that the number of edge state crossings is equal to the bulk index, which can be determined by counting the winding number of the eigenvalues of the transition matrix parametrized by $k \in S^1$, it suffices to prove the claim that at each crossing point, the winding number changes by ± 1. This turns a global problem into a local one. Hence, we focus on a particular crossing point (ε_F, k_*).

6. The transition matrices $T(k)$ and the matrices $L(z, k)$

From the way in which Ψ is defined, we see that the boundary values $\Psi_\pm(k)$ are given by

$$\Psi_\pm(k) = \widetilde{\Psi}_\pm(\varepsilon_F, k) M_\pm(k)^{-1}.$$

The transition matrices $T(k)$ relating the boundary values differ from the identity only inside the interval I, where they are given by

$$T(k) = M_-(k) M_+(k)^{-1}.$$

By the construction of this particular section, we have

$$\widetilde{\Psi}_0(z,k) = M_\pm(k) \quad \text{for } (z,k) \in (\partial D)_\pm,$$

and since $\widetilde{\Psi}$ is defined near the crossing point $(\varepsilon_{\mathrm{F}}, k_*)$, we can apply step 3 to obtain

$$L(z,k) = -\widetilde{\Psi}_1^\dagger(\overline{z},k)A(k)\widetilde{\Psi}_0(z,k) = -A(k)\widetilde{\Psi}_0(z,k).$$

From the above steps we hence see that the transition matrices $T(k)$ and the $L(z,k)$ are related by

$$T(k) = A(k)^{-1}L_-(k)L_+(k)^{-1}A(k),$$

where $L_\pm(k) = L(z,k)$ for $(z,k) \in (\partial D)_\pm$. Since we conjugate by $A(k)$ in the above equation, the eigenvalues of $T(k)$ are equal to the eigenvalues of $L_-(k)L_+(k)^{-1}$. This establishes the link between the transition matrix $T(k)$, which encodes the bulk index, and the matrices $L(z,k)$, which encode the edge states as we have seen in step 3.

8. Eigenvalues of $T(k)$ and $L(z,k)$

The final step of the proof is the most technical, and it consists of relating the winding number of the eigenvalues of $T(k)$ to $L(z,k)$ near the crossing point $(\varepsilon_{\mathrm{F}}, k_*)$ as k moves past k_*. By working to first order and considering an arbitrarily small disk D, it can be shown that as k passes k_*, one of the eigenvalues of $T(k)$ changes its winding number by ± 1 whereas the winding number of the other eigenvalues remains zero. This proves the claim stated in step 5, and therefore this final step proves that $\mathcal{I} = \mathcal{I}^\sharp$, which is the bulk-edge correspondence.

We conclude this section by remarking that the steps given here are by no means complete and that many details have been omitted to keep this section moderate in size. The full proof is given in [19], and this section could provide a guide for a more detailed study of the proof.

7 Discussion

In this project we have studied topological insulators in the framework of band theory using tight-binding models. We have constructed a Slater-Koster tight-binding model for Bi_2Se_3, with parameters obtained from first-principles DFT computations [17]. From this model we obtained a band structure that agrees with the literature [15]. Subsequently, we have computed the $\mathbb{Z}_2$-index of Bi_2Se_3 using two different methods, showing that they arrive at the same result: the $\mathbb{Z}_2$-index is non-trivial, implying through the bulk-boundary correspondence that Bi_2Se_3 is a topological insulator. We have computed and analyzed the density of states at the surface of Bi_2Se_3, showing that there are topologically protected surface states. The resulting surface density of states has been compared to experiments, showing qualitative agreement in the results: both show the presence of a Dirac cone on the surface of Bi_2Se_3. This confirms that the phenomena emerging in topological insulators can be understood using band theory. As Hasan and Kane wrote in [16]: "*It is remarkable that after more than 80 years, there are still treasures to be uncovered within band theory.*"

For further work, the tight-binding model that has been constructed for Bi_2Se_3 can be generalized to other topological insulators with the same crystal structure, such as bismuth telluride (Bi_2Te_3) and antimony telluride (Sb_2Te_3) [15]. The approach of surface Green's functions that has been used can potentially be extended to study the spin-properties of the topological surface states, which have not been investigated in this work. On the mathematical side, perhaps more direct proofs of the bulk-edge correspondence can be found, for instance based on scattering theory as suggested in [19].

References

[1] J. C. Slater and G. F. Koster, "Simplified LCAO method for the periodic potential problem," *Physical Review*, vol. 94, no. 6, pp. 1498–1524, 1954.

[2] J. M. Kosterlitz and D. J. Thouless, "Ordering, metastability and phase transitions in two-dimensional systems," *Journal of Physics C: Solid State Physics*, vol. 6, no. 7, pp. 1181–1203, 1973.

[3] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. Den Nijs, "Quantized Hall conductance in a two-dimensional periodic potential," *Physical Review Letters*, vol. 49, no. 6, pp. 405–408, 1982.

[4] F. D. M. Haldane, "Model for a quantum Hall effect without Landau levels: Condensed-matter realization of the "parity anomaly"," *Physical Review Letters*, vol. 61, no. 18, pp. 2015–2018, 1988.

[5] N. Ashcroft and N. Mermin, "Solid State Physics (brooks cole, 1976)," 1993.

[6] N. Marzari and D. Vanderbilt, "Maximally localized generalized Wannier functions for composite energy bands," *Physical Review B - Condensed Matter and Materials Physics*, vol. 56, no. 20, pp. 12 847–12 865, 1997.

[7] C. L. Kane and E. J. Mele, "$\mathbb{Z}_2$ topological order and the quantum spin hall effect," *Physical Review Letters*, vol. 95, no. 14, 2005.

[8] P. A. Khomyakov, G. Brocks, V. Karpan, M. Zwierzycki, and P. J. Kelly, "Conductance calculations for quantum wires and interfaces: Mode matching and Green's functions," *Physical Review B - Condensed Matter and Materials Physics*, vol. 72, no. 3, 2005.

[9] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, "Quantum spin Hall effect and topological phase transition in HgTe quantum wells," *Science*, vol. 314, no. 5806, pp. 1757–1761, 2006.

[10] L. Fu and C. L. Kane, "Time reversal polarization and a $\mathbb{Z}_2$ adiabatic spin pump," *Physical Review B - Condensed Matter and Materials Physics*, vol. 74, no. 19, 2006.

[11] L. Fu and C. L. Kane, "Topological insulators with inversion symmetry," *Physical Review B - Condensed Matter and Materials Physics*, vol. 76, no. 4, 2007.

[12] L. Fu, C. L. Kane, and E. J. Mele, "Topological insulators in three dimensions," *Physical Review Letters*, vol. 98, no. 10, 2007.

[13] R. Shankar, *Principles of Quantum Mechanics*, Second Edition. Springer, New York, 2008.

[14] 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 Physics*, vol. 5, no. 6, pp. 398–402, 2009.

[15] H. Zhang, C.-X. Liu, X.-L. Qi, X. Dai, Z. Fang, and S.-C. Zhang, "Topological insulators in Bi_2Se_3, Bi_2Te_3 and Sb_2Te_3 with a single Dirac cone on the surface," *Nature Physics*, vol. 5, no. 6, pp. 438–442, 2009.

[16] M. Z. Hasan and C. L. Kane, "Colloquium: Topological insulators," *Reviews of Modern Physics*, vol. 82, no. 4, pp. 3045–3067, 2010.

[17] K. Kobayashi, "Electron transmission through atomic steps of Bi_2Se_3 and Bi_2Te_3 surfaces," *Physical Review B - Condensed Matter and Materials Physics*, vol. 84, no. 20, 2011.

[18] A. A. Soluyanov and D. Vanderbilt, "Computing topological invariants without inversion symmetry," *Physical Review B - Condensed Matter and Materials Physics*, vol. 83, no. 23, 2011.

[19] G. M. Graf and M. Porta, "Bulk-edge correspondence for two-dimensional topological insulators," *Communications in Mathematical Physics*, vol. 324, no. 3, pp. 851–895, 2013.

[20] B. C. Hall, *Quantum Theory for Mathematicians*, ser. Graduate Texts in Mathematics. Springer, New York, 2013, vol. 267.

[21] A. Pertsova and C. M. Canali, "Probing the wavefunction of the surface states in Bi_2Se_3 topological insulator: A realistic tight-binding approach," *New Journal of Physics*, vol. 16, 2014.

[22] J. K. Asbóth, L. Oroszlány, and A. Pályi, *A short course on topological insulators*, ser. Lecture Notes in Physics. Springer, Cham, 2016, vol. 919.

[23] M. Farmanbar, T. Amlaki, and G. Brocks, "Green's function approach to edge states in transition metal dichalcogenides," *Physical Review B*, vol. 93, no. 20, 2016.

[24] E. Prodan and H. Schulz-Baldes, *Bulk and boundary invariants for complex topological insulators*, ser. Mathematical Physics Studies. Springer, 2016, From K-theory to physics.

[25] D. Carpentier, "Topology of bands in solids: From insulators to Dirac matter," Progress in Mathematical Physics, vol. 71, pp. 95–129, 2017.

[26] P. Woit, *Quantum Theory, Groups and Representations*. Springer, Cham, 2017.

A   Review of Quantum Mechanics

In this appendix we give a review of the principles of quantum mechanics that we need to in the main text of the thesis. This appendix will be brief and is not intended as a first introduction to the subject. An excellent introduction to the subject is [13]. For a more mathematical introduction, we refer the reader to [20] for an approach from the viewpoint of functional analysis and [26] for a treatment that is more focused on the algebra and representation theory that underlies quantum theory. One of the aims of this appendix is to bridge a gap between the different notations used in physical and mathematical literature.

Quantum mechanics emerged in the beginning of the 20th century, when it was realized that the physics at that time, now called classical physics, was not an adequate description of nature, especially at microscopic scales. The principles of quantum theory still form the basis for most microscopic theories today. In order to have a microscopic description of solid materials, we thus need to work in the framework of quantum mechanics.

A.1   Quantum Systems

A quantum-mechanical system consists of a complex separable Hilbert space $\mathcal{H}$ with a distinguished self-adjoint operator $H : \mathcal{H} \to \mathcal{H}$, called the Hamiltonian. We adopt the notational convention that is used throughout the physics literature, where we denote elements of $\mathcal{H}$ by symbols of the form

$$|\psi\rangle \in \mathcal{H},$$

which is called a *ket*. Through the Riesz representation theorem, we can isometrically identify $\mathcal{H}$ with its dual space $\mathcal{H}^*$ consisting of continuous linear functionals $\mathcal{H} \to \mathbb{C}$. In physics, this isomorphism is called Hermitian conjugation, and one writes the Riesz isomorphism as

$$\mathcal{H} \to \mathcal{H}^*$$
$$|\psi\rangle \mapsto \big(|\psi\rangle\big)^\dagger =: \langle\psi|.$$

The object on the right-hand side is called a *bra*. The inner product on $\mathcal{H}$ is then denoted by juxtaposition of a dual vector and a vector, i.e.

$$\langle\psi|\phi\rangle \in \mathbb{C}.$$

In the physics literature one also often encounters expressions of the form

$$P = |\psi\rangle \langle\phi|,$$

which is to be interpreted as an operator $\mathcal{H} \to \mathcal{H}$. This becomes clear if one considers the action of this operator on an element $|\chi\rangle \in \mathcal{H}$,

$$P|\chi\rangle = |\psi\rangle \langle\phi|\chi\rangle = \langle\phi|\chi\rangle |\psi\rangle \in \mathcal{H}.$$

A normalized element in $\mathcal{H}$ represents a state of the system[18]. The time evolution of the state of the system is generated by the Hamiltonian, according to the Schrödinger equation

$$i\hbar\partial_t |\psi(t)\rangle = H |\psi(t)\rangle.$$

---

[18]Strictly speaking, this works only for pure states. One can also consider mixed states, for which there is no direct correspondence between vectors in $\mathcal{H}$ and quantum states.

From this equation it follows that the time evolution of eigenvectors of H is particularly simple: if we consider a state $|\psi(0)\rangle$ satisfying

$$H |\psi(0)\rangle = E |\psi(0)\rangle ,$$

then Schrödinger equation is solved by

$$|\psi(t)\rangle = e^{-iEt/\hbar} |\psi(0)\rangle .$$

For this reason, one is often interested in the spectral decomposition of the Hamiltonian H. One is then led to diagonalize H by solving the eigenvalue equation, which in this context is called the time-independent Schrödinger equation.

A.2   Quantization and Second Quantization

Whereas classical observables are smooth functions on the phase space of the system, quantum observables are represented by operators on the underlying Hilbert space. The mathematical procedure of turning classical observables into quantum observables is called quantization.

Although we will not use quantum many-body systems or quantum field theory, it will be useful to briefly discuss the formalism of second quantization in a single-particle context. In this formalism one can consider quantum systems in which the number of particles is not fixed, and operators are expressed using so-called creation and annihilation operators. A creation operator $c_i^\dagger$ has the action of adding a particle in the state i to the state on which it acts, provided that this is allowed by symmetry considerations. Similarly, the annihilation operator c_i removes a particle from the state on which it acts, provided that there is a particle in this state. To make sense of these operators individually, one of course needs to work in a Hilbert space that admits states with a non-fixed number of particles. However, products of the form $c_i^\dagger c_j$ can be defined for single-particle Hilbert spaces. Consider a Hilbert space $\mathcal{H}$ spanned by an orthonormal basis $\{|i\rangle : i \in I\}$, where I is an index set. We define the operator

$$c_i^\dagger c_j : \mathcal{H} \to \mathcal{H}$$

by setting

$$c_i^\dagger c_j |k\rangle = \delta_{jk} |i\rangle$$

on the basis $\{|i\rangle : i \in I\}$ and extending linearly. Here δ_{ij} is the Kronecker delta. This operator has the interpretation of sending a particle in the state $|j\rangle$ to the state $|i\rangle$, whereas it sends all other basis vectors to 0. This is a particularly useful interpretation when one considers a Hamiltonian with terms that describe particles hopping between different states, as for example in tight-binding models. Note that if $A : \mathcal{H} \to \mathcal{H}$ is an operator with matrix elements

$$\langle i| A |j\rangle = A_{ij},$$

then one can reconstruct this operator from creation and annihilation operators by

$$A = \sum_{(i,j)\in I\times I} A_{ij} c_i^\dagger c_j, \tag{31}$$

since the right-hand side has the matrix elements

$$\langle k| \sum_{(i,j)\in I\times I} A_{ij} c_i^\dagger c_j |l\rangle = \sum_{(i,j)\in I\times I} A_{ij} \delta_{jl} \langle k|i\rangle = A_{kl}.$$

B Details of the Bi₂Se₃ Tight-Binding Model

In this appendix we elaborate further on the computation of the band structure of Bi$_2$Se$_3$. We first provide the details of the geometry of the crystal structure and our conventions for the unit cell. Then we discuss the way in which the tight-binding Hamiltonian in equation (17) is turned into a 40×40 matrix-valued function of $\mathbf{k}$ in the Brillouin zone.

B.1 Geometry of the Unit Cell

One of the ingredients of equation (17) is the relative vectors $\mathbf{r}_{ij}$ of all nearest neighbour and second nearest neighbour pairs of atoms in the crystal. For the convenience of the reader, the crystal structure and the unit cell are shown again below in figure 18. Based on this crystal structure and on the distances between the neighbouring atoms as given in [17], we can obtain the $\mathbf{r}_{ij}$ by reconstructing the coordinates of all atoms near a given unit cell. To this end, it is convenient to introduce a third basis $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$, along with the Cartesian basis and the basis of primitive lattice vectors. The vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ lie in the xy-plane and form the relative vectors between atoms inside the same single-atom layer as shown in figure 18. The third vector $\mathbf{a}_3$ is a vector in the z-direction with a length of 1 Å.
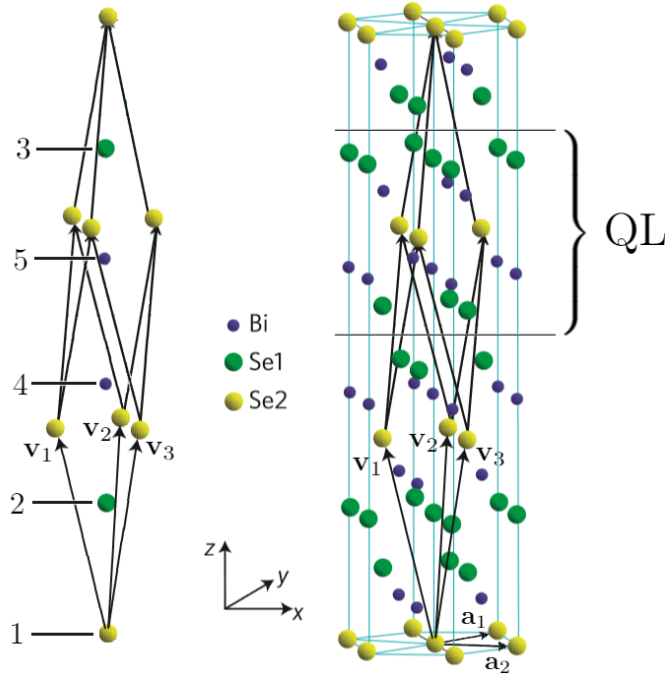
FIGURE 18: Crystal structure of Bi$_2$Se$_3$ with primitve lattice vectors $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ spanning the Bravais lattice as well as two vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ spanning the triangular lattice of the atomic layers parallel to the xy-plane. Se atoms at inequivalent positions are labelled by Se1 and Se2. Left: rhombohedral unit cell, showing our convention for the labels of the atomic sites in the unit cell. Right: full crystal structure in the neighbourhood of a unit cell. A quintuple layer (QL) is indicated. Adapted from Zhang et al. [15].

The tables below show the relative coordinates of the neighbouring atoms for each atom in the unit cell, expressed as $\mathbf{r} = a_1\mathbf{a}_1 + a_2\mathbf{a}_2 + a_3\mathbf{a}_3$. We distinguish intralayer neighbours (IL), nearest neighbours (NN) and second nearest neighbours (SN).

TABLE 2: Neighbourhood of atom 1 - Se(2).

IL	a_1	a_2	a_3	NN	a_1	a_2	a_3	SN	a_1	a_2	a_3
1 - Se(2)	1	0	0	1 - Bi	1/3	1/3	h_1	1 - Se(1)	2/3	-1/3	$h_1 + h_2$
2 - Se(2)	1	-1	0	2 - Bi	1/3	-2/3	h_1	2 - Se(1)	-1/3	-1/3	$h_1 + h_2$
3 - Se(2)	0	-1	0	3 - Bi	-2/3	1/3	h_1	3 - Se(1)	-1/3	2/3	$h_1 + h_2$
4 - Se(2)	-1	0	0	4 - Bi	2/3	-1/3	$-h_1$	4 - Se(1)	1/3	1/3	$-h_1 - h_2$
5 - Se(2)	-1	1	0	5 - Bi	-1/3	-1/3	$-h_1$	5 - Se(1)	1/3	-2/3	$-h_1 - h_2$
6 - Se(2)	0	1	0	6 - Bi	-1/3	2/3	$-h_1$	6 - Se(1)	-2/3	1/3	$-h_1 - h_2$

TABLE 3: Neighbourhood of atom 2 - Se(1).

IL	a_1	a_2	a_3	NN	a_1	a_2	a_3	SN	a_1	a_2	a_3
1 - Se(1)	1	0	0	1 - Bi	1/3	1/3	h_2	1 - Se(2)	2/3	-1/3	$h_1 + h_2$
2 - Se(1)	1	-1	0	2 - Bi	1/3	-2/3	h_2	2 - Se(2)	-1/3	-1/3	$h_1 + h_2$
3 - Se(1)	0	-1	0	3 - Bi	-2/3	1/3	h_2	3 - Se(2)	-1/3	2/3	$h_1 + h_2$
4 - Se(1)	-1	0	0	4 - Se(1)	2/3	-1/3	$-h_3$	4 - Bi	1/3	1/3	$-h_2 - h_3$
5 - Se(1)	-1	1	0	5 - Se(1)	-1/3	-1/3	$-h_3$	5 - Bi	1/3	-2/3	$-h_2 - h_3$
6 - Se(1)	0	1	0	6 - Se(1)	-1/3	2/3	$-h_3$	6 - Bi	-2/3	1/3	$-h_2 - h_3$

TABLE 4: Neighbourhood of atom 3 - Se(1).

IL	a_1	a_2	a_3	NN	a_1	a_2	a_3	SN	a_1	a_2	a_3
1 - Se(1)	1	0	0	1 - Se(1)	1/3	1/3	h_3	1 - Bi	2/3	-1/3	$h_2 + h_3$
2 - Se(1)	1	-1	0	2 - Se(1)	1/3	-2/3	h_3	2 - Bi	-1/3	-1/3	$h_2 + h_3$
3 - Se(1)	0	-1	0	3 - Se(1)	-2/3	1/3	h_3	3 - Bi	-1/3	2/3	$h_2 + h_3$
4 - Se(1)	-1	0	0	4 - Bi	2/3	-1/3	$-h_2$	4 - Se(2)	1/3	1/3	$-h_1 - h_2$
5 - Se(1)	-1	1	0	5 - Bi	-1/3	-1/3	$-h_2$	5 - Se(2)	1/3	-2/3	$-h_1 - h_2$
6 - Se(1)	0	1	0	6 - Bi	-1/3	2/3	$-h_2$	6 - Se(2)	-2/3	1/3	$-h_1 - h_2$

TABLE 5: Neighbourhood of atom 4 - Bi.

IL	a_1	a_2	a_3	NN	a_1	a_2	a_3	SN	a_1	a_2	a_3
1 - Bi	1	0	0	1 - Se(1)	1/3	1/3	h_2	1 - Se(1)	2/3	-1/3	$h_2 + h_3$
2 - Bi	1	-1	0	2 - Se(1)	1/3	-2/3	h_2	2 - Se(1)	-1/3	-1/3	$h_2 + h_3$
3 - Bi	0	-1	0	3 - Se(1)	-2/3	1/3	h_2	3 - Se(1)	-1/3	2/3	$h_2 + h_3$
4 - Bi	-1	0	0	4 - Se(2)	2/3	-1/3	$-h_1$	4 - Bi	1/3	1/3	$-2h_1$
5 - Bi	-1	1	0	5 - Se(2)	-1/3	-1/3	$-h_1$	5 - Bi	1/3	-2/3	$-2h_1$
6 - Bi	0	1	0	6 - Se(2)	-1/3	2/3	$-h_1$	6 - Bi	-2/3	1/3	$-2h_1$

TABLE 6: Neighbourhood of atom 5 - Bi.

IL	a_1	a_2	a_3	NN	a_1	a_2	a_3	SN	a_1	a_2	a_3
1 - Bi	1	0	0	1 - Se(2)	1/3	1/3	h_1	1 - Bi	2/3	-1/3	$2h_1$
2 - Bi	1	-1	0	2 - Se(2)	1/3	-2/3	h_1	2 - Bi	-1/3	-1/3	$2h_1$
3 - Bi	0	-1	0	3 - Se(2)	-2/3	1/3	h_1	3 - Bi	-1/3	2/3	$2h_1$
4 - Bi	-1	0	0	4 - Se(1)	2/3	-1/3	$-h_2$	4 - Se(1)	1/3	1/3	$-h_2 - h_3$
5 - Bi	-1	1	0	5 - Se(1)	-1/3	-1/3	$-h_2$	5 - Se(1)	1/3	-2/3	$-h_2 - h_3$
6 - Bi	0	1	0	6 - Se(1)	-1/3	2/3	$-h_2$	6 - Se(1)	-2/3	1/3	$-h_2 - h_3$

The distance constants in this table are

$$h_1 = 1.881 \text{ Å},$$
$$h_2 = 1.709 \text{ Å},$$
$$h_3 = 2.368 \text{ Å},$$

which are computed from the distances between the atoms as given in [17]. The hopping parameters $t_{ij}^{\alpha\alpha'}$ appearing in equation (17) are constructed from the so-called Slater-Koster parameters, obtained by Kobayashi [17]. For each pair of neighbouring atoms the Slater-Koster parameters for the p-orbitals, denoted by $V_{pp\sigma}$ and $V_{pp\pi}$, can be defined using two-center bond integrals of the p-orbitals on the considered atoms [1]. Consider a pair of atoms labelled by i and j. In the Cartesian basis, we can write

$$\mathbf{r}_{ij} = (l, m, n)|\mathbf{r}_{ij}|.$$

The numbers l, m, n are called the direction cosines of $\mathbf{r}_{ij}$. With this notation, the hopping parameters associated to i and j are constructed using [1]

$$t^{x,x} = l^2 V_{pp\sigma} + (1 - l^2)V_{pp\pi},$$
$$t^{x,y} = lm V_{pp\sigma} - lm V_{pp\pi},$$
$$t^{x,z} = ln V_{pp\sigma} - ln V_{pp\pi},$$
$$t^{y,y} = m^2 V_{pp\sigma} + (1 - m^2)V_{pp\pi},$$
$$t^{y,z} = mn V_{pp\sigma} - mn V_{pp\pi},$$
$$t^{y,x} = ml V_{pp\sigma} - ml V_{pp\pi},$$
$$t^{z,z} = n^2 V_{pp\sigma} + (1 - n^2)V_{pp\pi},$$
$$t^{z,x} = nl V_{pp\sigma} - nl V_{pp\pi},$$
$$t^{z,y} = nm V_{pp\sigma} - nm V_{pp\pi},$$

where $V_{pp\sigma}$ and $V_{pp\pi}$ are the Slater-Koster parameters associated to the pair (i, j) [1].

C Partial Bloch Hamiltonian of the BHZ model

In this appendix we derive the expression for the partial Bloch Hamiltonian of the BHZ model for k_y, which is given by

$$
H(k_y) = \frac{1}{2} \sum_{i_x} \left[|i_x + 1\rangle \langle i_x| \otimes s_0 \otimes \sigma_z + |i_x\rangle \langle i_x + 1| \otimes s_0 \otimes \sigma_z \right]
$$

$$
+ \frac{i}{2} \sum_{i_x} \left[|i_x + 1\rangle \langle i_x| \otimes s_z \otimes \sigma_x - |i_x\rangle \langle i_x + 1| \otimes s_z \otimes \sigma_x \right]
$$

$$
+ \sum_{i_x} |i_x\rangle \langle i_x| \otimes \left[s_x \otimes C + s_0 \otimes \left[(u + \cos k_y)\sigma_z + \sin k_y \sigma_y \right] \right].
$$

As a starting point, we consider the full BHZ Bloch Hamiltonian

$$
H_{\mathrm{BHZ}}(k_x, k_y) = s_0 \otimes \left[(u + \cos k_x + \cos k_y)\sigma_z + \sin k_y \sigma_y \right] + s_z \otimes \sin k_x \sigma_x + s_x \otimes C
$$

for a finite lattice $B = \mathbb{Z}_{N_x} \oplus \mathbb{Z}_{N_y}$ with periodic boundary conditions, with N_x sites in the x-direction and N_y sites in the y-direction. The full Bloch Hamiltonian given above is related to the partial Bloch Hamiltonian $H(k_y)$ by

$$
H(k_x, k_y) = \langle k_x| H(k_y) |k_x\rangle, \quad H(k_y) = \sum_{k_x} |k_x\rangle \langle k_x| \otimes H(k_x, k_y).
$$

To keep the derivation clear, we construct $H(k_y)$ term by term. Hence, we define

$$
H_{\mathrm{I}}(k_x, k_y) = s_0 \otimes \left[(u + \cos k_x + \cos k_y)\sigma_z + \sin k_y \sigma_y \right],
$$

$$
H_{\mathrm{II}}(k_x, k_y) = s_z \otimes \sin k_x \sigma_x,
$$

$$
H_{\mathrm{III}}(k_x, k_y) = s_x \otimes C.
$$

We use the conventional abbreviation +H.c. to indicate that the Hermitian conjugate of the preceding term is to be added. Using the definition of the plane wave $|k_x\rangle$,

$$
|k_x\rangle = \frac{1}{\sqrt{N_x}} \sum_{i_x} e^{ik_x i_x} |i_x\rangle,
$$

it follows that

$$
H_{\mathrm{I}}(k_y) = \sum_{k_x} |k_x\rangle \langle k_x| \otimes H_{\mathrm{I}}(k_x, k_y)
$$

$$
= \sum_{k_x} |k_x\rangle \langle k_x| \otimes \left[s_0 \otimes \left[(u + \frac{1}{2}(e^{ik_x} + e^{-ik_x}) + \cos k_y)\sigma_z + \sin k_y \sigma_y \right] \right]
$$

$$
= \sum_{k_x} |k_x\rangle \langle k_x| \otimes \left[s_0 \otimes \left[(u + \frac{1}{2}(e^{ik_x} + e^{-ik_x}) + \cos k_y)\sigma_z + \sin k_y \sigma_y \right] \right]
$$

$$
= \frac{1}{2N_x} \sum_{k_x} \sum_{i_x} \sum_{i_x'} e^{ik_x(i_x + 1 - i_x')} |i_x\rangle \langle i_x'| \otimes s_0 \otimes \sigma_z + \mathrm{H.c.}
$$

$$
+ \sum_{i_x} |i_x\rangle \langle i_x| \otimes s_0 \otimes \left[(u + \cos k_y)\sigma_z + \sin k_y \sigma_y \right]
$$

$$
= \frac{1}{2} \sum_{i_x} |i_x + 1\rangle \langle i_x| \otimes s_0 \otimes \sigma_z + \mathrm{H.c.}
$$

$$
+ \sum_{i_x} |i_x\rangle \langle i_x| \otimes s_0 \otimes \left[(u + \cos k_y)\sigma_z + \sin k_y \sigma_y \right],
$$

where we have used the fact that the identity on $\ell^2(B)$ can be expressed as

$$1 = \sum_{k_x} |k_x\rangle \langle k_x| = \sum_{i_x} |i_x\rangle \langle i_x|.$$

For the second term, we have

$$
\begin{aligned}
H_{\mathrm{II}}(k_y) &= \sum_{k_x} |k_x\rangle \langle k_x| \otimes H_{\mathrm{II}}(k_x, k_y) \\
&= \sum_{k_x} |k_x\rangle \langle k_x| \otimes \Big[-s_z \otimes \frac{i}{2}(e^{ik_x} - e^{-ik_x})\sigma_x \Big] \\
&= \frac{1}{N_x} \sum_{k_x} \sum_{i_x} \sum_{i_x'} e^{ik_x(i_x - i_x')} |i_x\rangle \langle i_x'| \otimes \Big[-s_z \otimes \frac{i}{2}(e^{ik_x} - e^{-ik_x})\sigma_x \Big] \\
&= \frac{-i}{2N_x} \sum_{k_x} \sum_{i_x} \sum_{i_x'} e^{ik_x(i_x + 1 - i_x')} |i_x\rangle \langle i_x'| \otimes s_z \otimes \sigma_x \\
&\quad + \frac{i}{2N_x} \sum_{k_x} \sum_{i_x} \sum_{i_x'} e^{ik_x(i_x - 1 - i_x')} |i_x\rangle \langle i_x'| \otimes s_z \otimes \sigma_x \\
&= \frac{i}{2} \sum_{i_x} |i_x + 1\rangle \langle i_x| \otimes s_z \otimes \sigma_x + \mathrm{H.c.}
\end{aligned}
$$

The transformation of the third term is simply given by

$$
\begin{aligned}
H_{\mathrm{III}}(k_y) &= \sum_{k_x} |k_x\rangle \langle k_x| \otimes H_{\mathrm{III}}(k_x, k_y) \\
&= \sum_{k_x} |k_x\rangle \langle k_x| \otimes s_x \otimes C \\
&= \sum_{i_x} |i_x\rangle \langle i_x| \otimes s_x \otimes C.
\end{aligned}
$$

Adding all three terms together, we finally obtain

$$
\begin{aligned}
H(k_y) &= \frac{1}{2} \sum_{i_x} \Big[|i_x + 1\rangle \langle i_x| \otimes s_0 \otimes \sigma_z + |i_x\rangle \langle i_x + 1| \otimes s_0 \otimes \sigma_z \Big] \\
&\quad + \frac{i}{2} \sum_{i_x} \Big[|i_x + 1\rangle \langle i_x| \otimes s_z \otimes \sigma_x - |i_x\rangle \langle i_x + 1| \otimes s_z \otimes \sigma_x \Big] \\
&\quad + \sum_{i_x} |i_x\rangle \langle i_x| \otimes \Big[s_x \otimes C + s_0 \otimes \big[(u + \cos k_y)\sigma_z + \sin k_y \sigma_y \big] \Big],
\end{aligned}
$$

which is the expression that we wanted to derive.