Band structures of topological crystalline insulators

Bandstrukturer för topologiska kristallina isolatorer

Elisabet Edvardsson

Faculty of Health, Science and Technology

Degree Project for Master of Science in Engineering, Engineering Physics

30 ECTS

Supervisor: Jürgen Fuchs

Examiner: Lars Johansson

January 2018

Abstract

Topological insulators and topological crystalline insulators are materials that have a bulk band structure that is gapped, but that also have toplogically protected non-gapped surface states. This implies that the bulk is insulating, but that the material can conduct electricity on some of its surfaces. The robustness of these surface states is a consequence of time-reversal symmetry, possibly in combination with invariance under other symmetries, like that of the crystal itself. In this thesis we review some of the basic theory for such materials. In particular we discuss how topological invariants can be derived for some specific systems. We then move on to do band structure calculations using the tight-binding method, with the aim to see the topologically protected surface states in a topological crystalline insulator. These calculations require the diagonalization of block tridiagonal matrices. We finish the thesis by studying the properties of such matrices in more detail and derive some results regarding the distribution and convergence of their eigenvalues.

Acknowledgements

First of all I would like to thank my supervisor Prof. Jürgen Fuchs for introducing me to the interesting subject of topological phases and for always taking the time to discuss my questions – even when there was no time for it.

Thank you also to Prof. Ryszard Buczko, Prof. Lars Johansson and Dr. Thijs Jan Holleboom for discussing and answering my questions regarding the slab method used in the tight-binding calculations.

A special thank you to Dr. Eva Mossberg, a wonderful friend and mentor during my years at Karlstad University, for many interesting conversations about eigenvalues and for helping me find a bug in the MATLAB version I was using.

Finally, I would like to thank my parents for their constant support and for telling me when it was time to take a break.

Contents

1 Introduction

The notion of phases in materials is very common. Traditionally, when talking about phases one refers to the material being either in a solid, liquid or gaseous state. This, however, is a very rough division of materials into different classes, since for example, many common materials, such as steel or ice, can in fact exist is several different forms. A finer division of materials arises from what in [1] is called the *principle of emergence*, which states that it is the organization of the particles in a material that determines the properties of the material. The problem is thus to describe how particles are ordered in materials, and to use these different orders to describe different phases. What one needs to do first is to define when two orderings of particles should be considered equivalent and in [1] they have the following:

Two states that can be connected to each other without any phase transitions are equivalent.

When we say that we connect two states, we mean that we start in one state and deform the system smoothly in some way, e.g. by changing the temperature, until we end up in the other state. Quantum mechanically, this means that we begin with a Hamiltonian which depends on a set of parameters, and then we smoothly change some of the parameters so that the Hamiltonian changes. By a *phase transition*, one means that there is at least one local quantity that does not change smoothly under the deformation of the system. [1]

One theory that describes phases and phase transitions and when states are equivalent, was developed by Landau [2]. In this theory, the main feature is symmetries, and we say that different phases have different symmetries and that phases change when symmetries are broken. This means that a phase transition is a transition that changes the symmetry of the material.[1]

Landau's theory was successful, and one consequence of it, for example, is that we can classify all three-dimensional crystal structures. For a long time it was believed that this description of order in materials was complete. This, however, has in the last 30 years turned out not to be the case. The two main features that led to this conclusion was the discovery of the fractional quantum Hall effect by Tsui and Störmer in 1982 [3], and that of high T_c superconductors by Bednorz and Müller in 1986 [4]. In the case of the fractional quantum Hall effect, the situation is such that there are different fractional quantum Hall states that have the same symmetry. Thus something is lacking in the classification of phases in terms of symmetry, and the need to describe other kinds of order in materials arose.

These new kinds of orders that can arise in materials are called *topological phases*, and are thus orders that are not described by symmetry breaking alone. It turns out that one feature that can be used to characterize these topological phases in thermodynamical systems is topology-dependent and topologically robust degeneracies in the ground state when considering the thermodynamic limit. It turns out that many of these properties can be described in terms of suitable topological invariants, which can take different quantized values for different phases, which explains why it is called topological phases.

In this thesis, we will focus on a particular class of such phases, which are called topological insulators and topological crystalline insulators. These are materials that are characterized by certain surface or edge states that exist as a consequence of a non-trivial topology of the wave functions of the bulk material. [5] The difference between a trivial insulator and a topological

insulator lies in the existence of such states. All insulators have a bandgap in their bulk band structure. This is simply a gap between the conduction band and the valence band. In a trivial insulator with a surface, we can have surface states that reach into the bandgap, giving the material conducting properties. These conducting properties, however, can be removed by changing the Fermi level of the material, so that it once again ends up in a bandgap. This means that in a trivial insulator, the Fermi level must cross each energy band an even number of times. In a topological insulator, however, there are energy bands that cross the Fermi level an odd number of times, meaning that there is no way that the Fermi level can be moved in a way so that it does not cross any energy bands. These surface states are robust and have the nice property that they are insensitive to e.g. contamination of the surface. [6, Ch. 2.1]

The most widely studied topological insulators are those for which the topological surface states are protected by time-reversal symmetry. In those systems the surface states have a Dirac dispersion, meaning that there are linear crossings between energy states. In topological crystalline insulators the situation is somewhat different. Here time-reversal symmetry alone is not enough to guarantee topologically protected states, instead one has to consider time-reversal symmetry in combination with symmetries of the crystal structure of the material itself.

As already mentioned, the theory of topological phases is relatively new. Actually, the Nobel Prize in physics was awarded "for theoretical discoveries of topological phase transitions and topological phases of matter" [7] in 2016 to Thouless, Haldane and Kosterlitz. The novelty of the materials means that they are not in use yet, but their properties, like the combination of spin polarization and large robustness of the surface states [6, Ch. 2.3], make them promising for applications in e.g. electronics and sensors.

One goal of this thesis is to describe the basics of topological insulators and topological crystalline insulators from a theoretical point of view. The goal is to give an overview of the theoretical aspects of the materials. In addition then shift the focus to band structure calculations using the tight-binding approximation. The structure of the thesis is as follows. In Section 3, we review the basics of topological insulators. We follow the description given in [5], and fill in the details, of the topological invariants associated with these kinds of materials and see why and how they arise. We relate these invariants to the band structure and see how the surface states differ in the different phases. In Section 4, we move on to review topological crystalline insulators and see how these differ from the previously described topological insulators. We follow the approach of [8].

In order to be able to show that a material is a topological crystalline insulator, one has to be able to calculate the band structure of the material. It is impossible to do this exactly, and thus in Section 5 we review the foundation of the tight-binding approximation and the Slater-Koster rules, which is described in [9]. In Section 6 we continue to some examples of how to use the Slater-Koster rules. We show how these rules are used to calculate the band structure in case of the material and in a general tetragonal crystal structure. In the first case we evaluate the bulk band structure using s, p and d orbitals, while in the latter case we use only p_x and p_y orbitals. In the case of the tetragonal lattice, we also perform tight-binding calculations for a slab in order to find surface states that cross each other and thus indicate that we are actually dealing with a topological crystalline insulator. When doing tight-binding calculations for a slab-geometry, one ends up with the problem of finding eigenvalues of a block-tridagonal matrix. Diagonalization of matrices is a computationally expensive problem, and thus we spend Section 7 on studying the properties of the eigenvalues of

these matrices. We provide limits on intervals in which these eigenvalues must lie. The results give a mathematical argument for why these kinds of matrices give rise to band structures in materials. Also, we do further examinations of these matrices and discuss the convergence of the eigenvalues (and thus the convergence of the band structure obtained in slab-geometry calculations) as the size of the matrix increases.

2 Some necessary physical concepts

In this section we will provide a background to some of the physical concepts that will be of importance in this thesis. We will among other things describe the adiabatic approximation, parity operators and the time-reversal operator.

2.1 Adiabatic systems

In many cases we will be interested in physical systems that vary slowly with time. These systems are called adiabatic, and in this section we will give a more accurate description of them that is based on the information in [10] and [11].

Suppose that we have a system with a Hamiltonian that depends on a set of parameters. The energy eigenvalues of the Hamiltonian will naturally depend on those parameters. Now, if these parameters vary slowly with time, the energy eigenvalues should not change their order. By *slowly varying* one usually means that they vary on a time scale that is much larger than $2\pi/\omega_{ab} \propto 1/E_{ab}$ for some difference in energy eigenvalues E_{ab}, where ω_{ab} is the frequency of the system. Such a change in parameters is called *adiabatic*.

An important result is the *adiabatic theorem*. It states the following [11]:

Suppose we have a time-dependent Hamiltonian. Then the eigenfunctions and eigenvalues of the system are time-dependent, giving us the equation

$$H(t)\psi_n(t) = E_n(t)\psi_n(t), \tag{2.1}$$

where the eigenfunctions at each instant of time are orthonormal to each other, i.e.

$$\langle \psi_n(t) | \psi_m(t) \rangle = \delta_{nm}. \tag{2.2}$$

Also, they form a complete set of basis functions, so we can express the solution to the general Schrödinger equation,

$$\mathrm{i}\hbar \frac{\partial}{\partial t}\Psi(t) = H(t)\Psi(t), \tag{2.3}$$

as a linear combination of the eigenfunctions in the following way

$$\Psi(t) = \sum_n \Psi_n(t) = \sum_n c_n(t)\psi_n(t)e^{\mathrm{i}\theta_n(t)}, \tag{2.4}$$

where

$$\theta_n(t) = -\frac{1}{\hbar}\int_0^t E_n(t')dt'. \tag{2.5}$$

Now, by inserting equation (2.4) into equation (2.3), and assuming non-degenerate energies, one can show that

$$\dot{c}_m(t) = -c_m \langle \psi_m | \dot{\psi}_m \rangle - \sum_{n \neq m} c_n \frac{\langle \psi_m | \dot{H} | \psi_n \rangle}{E_n - E_m} \exp\left[-\frac{\mathrm{i}}{\hbar} \int_0^t \left(E_n(t') - E_m(t') \right) dt' \right]. \tag{2.6}$$

The adiabatic approximation is now to assume that $\dot{H}$ is very small, in a sense that one can neglect the sum, thus leaving us with

$$\dot{c}_m(t) = -c_m \langle \psi_m | \dot{\psi}_m \rangle . \tag{2.7}$$

This is a differential equation with the solution

$$c_m(t) = c_m(0) \exp\left[\mathrm{i} \gamma_m(t) \right], \tag{2.8}$$

where

$$\gamma_m(t) = \mathrm{i} \int_0^t \langle \psi_m(t') | \frac{\partial}{\partial t'} \psi_m(t') \rangle \, dt'. \tag{2.9}$$

Now, if we assume that the system starts in the nth eigenstate at $t = 0$, then we have $c_n(0) = 1$ and $c_m(0) = 0$ for $m \neq n$. This means that

$$\Psi(t) = \Psi_n(t) = \exp\left[\mathrm{i} \theta_n(t) \right] \exp\left[\mathrm{i} \gamma_n(t) \right] \psi_n(t). \tag{2.10}$$

So the particle will remain in the nth eigenstate of the time evolving Hamiltonian, the only difference being some phase factors.

2.2 The Heisenberg equation of motion

The Heisenberg equation of motion is an equation that describes the time-evolution of an operator in the Heisenberg picture. We will give a brief review of it and follow the description in [10].

Let H be the Hamiltonian of a system and let A^S be an observable in the Schrödinger picture. Using this we define the corresponding observable in the Heisenberg picture by

$$A^H(t) := \mathcal{U}^\dagger(t) A^S \mathcal{U}(t), \tag{2.11}$$

where $\mathcal{U}$ is the time-evolution operator given by

$$\mathcal{U}(t) = \exp\left(\frac{-\mathrm{i} H t}{\hbar} \right). \tag{2.12}$$

By differentiating equation (2.11), we obtain

$$\frac{dA^H}{dt} = \frac{\partial \mathcal{U}^\dagger}{\partial t} A^S \mathcal{U} + \mathcal{U}^\dagger A^S \frac{\partial \mathcal{U}}{\partial t}. \tag{2.13}$$

From equation (2.12) we have

$$\frac{\partial \mathcal{U}}{\partial t} = \frac{1}{\mathrm{i}\hbar} H \mathcal{U}, \tag{2.14}$$

and

$$\frac{\partial \mathcal{U}^\dagger}{\partial t} = -\frac{1}{\mathrm{i}\hbar}\mathcal{U}^\dagger H. \tag{2.15}$$

Inserting this into equation (2.13), we get

$$\frac{dA^H}{dt} = -\frac{1}{\mathrm{i}\hbar}\mathcal{U}^\dagger H\mathcal{U}\mathcal{U}^\dagger A^S\mathcal{U} + \frac{1}{\mathrm{i}\hbar}\mathcal{U}^\dagger A^S\mathcal{U}\mathcal{U}^\dagger H\mathcal{U} = \frac{1}{\mathrm{i}\hbar}\left[A^H, \mathcal{U}^\dagger H\mathcal{U}\right]. \tag{2.16}$$

Because of equation (2.12), we see that H and $\mathcal{U}$ commute. This gives us

$$\frac{dA^H}{dt} = \frac{1}{\mathrm{i}\hbar}\left[A^H, H\right]. \tag{2.17}$$

This equation is known as the *Heisenberg equation of motion*.

2.3 The time-reversal operator

The time-reversal operator is central in the description of topological insulators, and we will here give a review of the most important properties based on [10] and [5].

Before we define the time-reversal operator, we will make some general notes about symmetry operators.

Definition 2.1. Let $|\alpha\rangle$ and $|\beta\rangle$ be two states. A *unitary* operator is a linear operator, U, that satisfies

$$\langle\tilde{\beta}|\tilde{\alpha}\rangle = \langle\beta|\alpha\rangle, \tag{2.18}$$

where $|\tilde{\alpha}\rangle = U|\alpha\rangle$ and $|\tilde{\beta}\rangle = U|\beta\rangle$.

Now, we will not only be interested in linear operators. In fact, it turns out that also *anti-linear* operators are useful. Such an operator θ satisfies the following:

$$\theta(c_1|\alpha\rangle + c_2|\beta\rangle) = c_1^*\theta|\alpha\rangle + c_2^*\theta|\beta\rangle. \tag{2.19}$$

Using this, we make the following definition:

Definition 2.2. An operator θ is *anti-unitary* if it is anti-linear and satisfies

$$\langle\tilde{\beta}|\tilde{\alpha}\rangle = \langle\beta|\alpha\rangle^*. \tag{2.20}$$

where $|\tilde{\alpha}\rangle = \theta|\alpha\rangle$ and $|\tilde{\beta}\rangle = \theta|\beta\rangle$.

It can be shown that an anti-unitary operator θ always can be written as

$$\theta = UK, \tag{2.21}$$

where U is a unitary operator and K is a complex conjugate operator. Some care is needed to be taken when writing θ in this way, since complex conjugation is not invariant under change-of-basis.

Now, we are specifically interested in the *time-reversal operator*. This is an anti-unitary operator that we denote by Θ. What follows is a discussion on the behaviour of operators under time-reversal.

Let $|\alpha\rangle$ and $|\beta\rangle$ be states in some system, let A be a linear operator, and define

$$|\tilde{\alpha}\rangle := \Theta\,|\alpha\rangle\,, \quad |\tilde{\beta}\rangle := \Theta\,|\beta\rangle\,. \tag{2.22}$$

Further, let $|\gamma\rangle = A^\dagger\,|\beta\rangle$. This gives us

$$\begin{aligned}\langle\beta|A|\alpha\rangle = \langle\gamma|\alpha\rangle = \langle\tilde{\alpha}|\tilde{\gamma}\rangle = \langle\tilde{\alpha}|\Theta A^\dagger|\beta\rangle \\ = \langle\tilde{\alpha}|\Theta A^\dagger\Theta^{-1}\Theta|\beta\rangle = \langle\tilde{\alpha}|\Theta A^\dagger\Theta^{-1}|\tilde{\beta}\rangle\,.\end{aligned} \tag{2.23}$$

In particular, if A is *Hermitian*, we have

$$\langle\beta|A|\alpha\rangle = \langle\tilde{\alpha}|\Theta A\Theta^{-1}|\tilde{\beta}\rangle\,. \tag{2.24}$$

We also note that if A is the identity operator, we have

$$\langle\alpha|\beta\rangle = \langle\Theta\beta|\Theta\alpha\rangle\,. \tag{2.25}$$

We say that an observable is *even* or *odd* under time-reversal depending on the sign in

$$\Theta A\Theta^{-1} = \pm A. \tag{2.26}$$

This means that

$$\langle\beta|A|\alpha\rangle = \pm\,\langle\tilde{\beta}|A|\tilde{\alpha}\rangle^*\,. \tag{2.27}$$

Letting $\alpha = \beta$, we get information about the expectation value under time-reversal, namely

$$\langle\alpha|A|\alpha\rangle = \pm\,\langle\tilde{\alpha}|A|\tilde{\alpha}\rangle\,. \tag{2.28}$$

In particular, it is clear that the expectation value of the momentum operator should change sign under time reversal, i.e.

$$\langle\alpha|\mathbf{p}|\alpha\rangle = -\,\langle\tilde{\alpha}|\mathbf{p}|\tilde{\alpha}\rangle\,, \tag{2.29}$$

which means that

$$\Theta\mathbf{p}\Theta^{-1} = -\mathbf{p}. \tag{2.30}$$

Similarly, the expectation value of the position operator should be unchanged under time reversal, giving us

$$\langle\alpha|\mathbf{x}|\alpha\rangle = \langle\tilde{\alpha}|\mathbf{x}|\tilde{\alpha}\rangle\,, \tag{2.31}$$

and

$$\Theta\mathbf{x}\Theta^{-1} = \mathbf{x}. \tag{2.32}$$

Now, it is also important to know how the wave function changes under time-reversal. One can show that

$$\Theta\psi(\mathbf{p}) = \psi^*(-\mathbf{p}). \tag{2.33}$$

It turns out that Θ behaves differently in systems with different spin. In systems with half-integer spin we have $\Theta^2 = -1$, while in systems with integer spin we have $\Theta^2 = 1$. In particular, we will

be interested in the case of spin 1/2 particles. Here Θ takes the form $\Theta = -\mathrm{i}s_y K$, where K is the complex conjugate operator and s_i denotes the spin operator given by Pauli matrices. We note that

$$\Theta^2 = (-\mathrm{i}s_y K)^2 = (s_y K)^2. \tag{2.34}$$

Since K commutes with s_i and $s_i^2 = -I$, we see that

$$\Theta^2 = -I. \tag{2.35}$$

2.4 The parity operator

Another useful symmetry operator is the parity operator. A parity operation can either be applied to the coordinate system or to the states themselves. Applying it to the coordinate system, amounts to changing the the system from a right-handed to a left-handed coordinate system. We will, however, be interested in the application of the parity operator to states. This is defined in the following way:

Definition 2.3. The *parity operator*, denoted by π, is a unitary operator which acts on any state $|\alpha\rangle$ such that the expectation value of $\mathbf{x}$ changes in the following way:

$$\langle\alpha|\pi^\dagger \mathbf{x}\pi|\alpha\rangle = -\langle\alpha|\mathbf{x}|\alpha\rangle\,. \tag{2.36}$$

One immediately sees that this is true if

$$\pi^\dagger \mathbf{x}\pi = -\mathbf{x}. \tag{2.37}$$

That π is unitary means

$$\pi^{-1} = \pi^\dagger, \tag{2.38}$$

which, together with equation (2.37) implies that π and $\mathbf{x}$ anti-commute, i.e.

$$\{\pi, \mathbf{x}\} = 0. \tag{2.39}$$

Now, let $|\mathbf{x}\rangle$ be an eigenstate of $\mathbf{x}$. Then we have

$$\pi\,|\mathbf{x}\rangle = |-\mathbf{x}\rangle\,. \tag{2.40}$$

From this it follows that

$$\pi^2\,|\mathbf{x}\rangle = |\mathbf{x}\rangle\,, \tag{2.41}$$

and thus π has eigenvalues ± 1.

The momentum operator behaves similarly under space inversion as the position operator, namely

$$\{\pi, \mathbf{p}\} = 0, \tag{2.42}$$

and

$$\pi^\dagger \mathbf{p}\pi = -\mathbf{p}. \tag{2.43}$$

2.5 Surface states

When studying topological insulators, one is interested in studying surface states of materials, i.e. electrons that are close to the surface. These electrons should have different properties compared to electrons existing in the bulk, since they do not have atoms on all sides, like those in the bulk do.

A surface state is described by its energy E and a two-dimensional wave vector (k_x, k_y) which is parallel to the surface [12, ch. 6.2.1], while bulk states are described by their energy and a three-dimensional wave-vector. In order to study bulk and surface states simultaneously, one has to project the bulk states onto the plane $E(k_x, k_y)$. The surface states in this description are characterized by not being degenerate with the bulk states, which means that they are found in the gap of the projected bulk band structure.

3 Topological insulators

In this section we will review the basic properties of topological insulators. In [13] two insulators are defined to be equivalent in the following way:

Definition 3.1. Two insulators are *topologically equivalent* if the Hamiltonians describing their band structures can be smoothly deformed into each other without closing the energy gap.

In practice this means the following: Suppose we have two systems. Now, start with one of them and smoothly change one or more of the parameters of the system. If we in this way can go from the first system to the other, while keeping the energy gap open, we say that the systems are topologically equivalent.

Our main reference in this section will be [5], and we will follow the presentation there, but fill in the details along the way.

3.1 The Hall effect

One can say that the first topological insulators that were discovered were the quantum Hall systems. For historical reasons we will thus start with a short summary of these systems and their properties.

The geometry of the Hall effect is shown in Figure 1. We have a two-dimensional system with a strong magnetic field $\mathbf{B} = (0, 0, B)$ in the z-direction and an electric field $\mathbf{E} = (E, 0, 0)$ in the x-direction.

3.1.1 The classical case

In the classical case the electric field creates a current $\mathbf{j} = \sigma\mathbf{E}$, where σ is the conductivity of the material. The magnetic field will exert a force on the electrons, so there will be a current in the

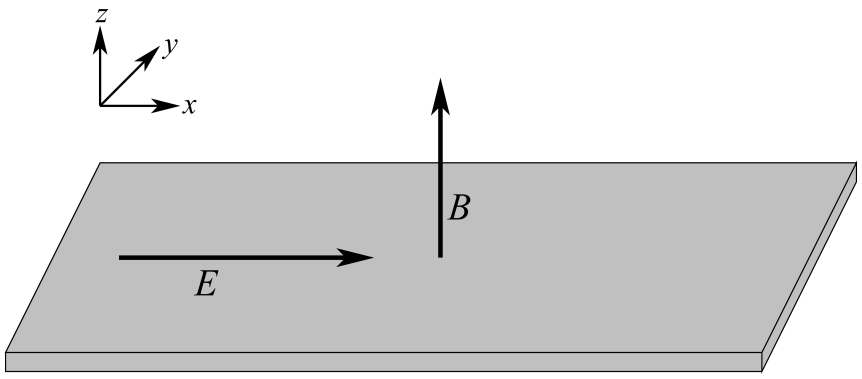

Figure 1: Geometry of the Hall experiment.

y-direction, which creates an electric field E_y in the y-direction which cancels the current. The transverse electric field is given by [14]

$$E_y = \sigma R_H E_x B_z, \tag{3.1}$$

where R_H is the Hall coefficient, given by

$$R_H = -\frac{1}{ne}, \tag{3.2}$$

where n is the electron density and e is the fundamental charge.

3.1.2 The quantum Hall effect

It turns out that when we have extremely low temperatures and strong magnetic fields, we will get quite a different phenomenon. This is called the quantum Hall effect. There are some different variants of this phenomenon, but we will discuss the integer quantum Hall effect, which was originally observed by von Klitzing, Dorda and Pepper in 1980 [14].

In this case the geometry of the system is the same as in the classical case. The difference however, is that when the temperature is of the order of a few Kelvin and the magnetic field is of a few Tesla, the Hall conductance is quantized according to

$$\sigma_{xy} = \nu \frac{e^2}{h}, \tag{3.3}$$

where h is Planck's constant.

The explanation for the quantized conductance lies in what happens to the electrons in the material in high magnetic fields. This is described in [15]. In a somewhat classical description, we can see that the electrons in the material will begin to make circular motions when in a strong magnetic field. If the width of the slab is large enough, this means that the electrons within the system will be localized. The electrons close to the edge, however, will start to move along the edge, so we get non-interacting edge channels moving in opposite directions at each edge of the material. This means that back-scattering in the material is suppressed close to integer filling factors.

3.1.3 The Berry phase

In order to understand a more technical description of how the integer quantum Hall system is a topological insulator, we first discuss the Berry phase. Let $\mathbf{R}(t)$ be a set of time-dependent parameters considered as a vector in parameter space. Now we consider a Hamiltonian specified by the parameters $\mathbf{R}(t)$, and denote it by $H[\mathbf{R}(t)]$. We also denote its nth eigenstate by $|n, \mathbf{R}(t)\rangle$, which gives us the following Schrödinger equation for the system:

$$H[\mathbf{R}(t)]\,|n, \mathbf{R}(t)\rangle = E_n[\mathbf{R}(t)]\,|n, \mathbf{R}(t)\rangle\,. \tag{3.4}$$

Now, assume that $\mathbf{R}(t)$ changes adiabatically, as described in Section 2.1, from $\mathbf{R}(t=0)$. If the the system starts in the nth state $|n, \mathbf{R}(t)\rangle$ (this notation corresponds to the wave function $\psi_n(t)$ in Section 2.1) we get the following time evolution for the system:

$$H[\mathbf{R}(t)]\,|n, t\rangle = \mathrm{i}\hbar\frac{\partial}{\partial t}\,|n, t\rangle\,, \tag{3.5}$$

where $|n, t\rangle$ corresponds to $\Psi_n(t)$ in Section 2.1. Now we can write, as is shown in [11],

$$\frac{\partial}{\partial t}\,|n, \mathbf{R}(t)\rangle = \dot{\mathbf{R}}(t)\nabla_{\mathbf{R}}\,|n, \mathbf{R}(t)\rangle\,. \tag{3.6}$$

This gives us, using equations (2.9) and (2.10), the following expression for the state at time t:

$$|n, t\rangle = \exp\left(\frac{\mathrm{i}}{\hbar}\int_0^t dt' L_n\left[\mathbf{R}(t')\right]\right)|n, \mathbf{R}(t)\rangle\,, \tag{3.7}$$

where

$$L_n[\mathbf{R}(t)] = \mathrm{i}\hbar\dot{\mathbf{R}}(t)\,\langle n, \mathbf{R}(t)|\nabla_R|n, \mathbf{R}(t)\rangle - E_n[\mathbf{R}(t)]\,. \tag{3.8}$$

Or, written differently,

$$|n, t\rangle = \exp\left(-\int_0^t dt'\dot{\mathbf{R}}(t')\,\langle n, \mathbf{R}(t')|\nabla_R|n, \mathbf{R}(t')\rangle\right)|n, \mathbf{R}(t)\rangle \times \exp\left(\frac{\mathrm{i}}{\hbar}\int_0^t dt' E_n\left[\mathbf{R}(t')\right]\right). \tag{3.9}$$

The first exponential term represents the non-trivial effect of the quantum-mechanical phase accumulated during the time evolution, and the last one is a trivial one called the dynamical term.

Now we consider the case when $\mathbf{R}$ moves on a closed loop C and returns from its original value $\mathbf{R}(t=0)$ at time $t=T$, so $\mathbf{R}(0) = \mathbf{R}(T)$. For such a loop, C, the *Berry phase*, $\gamma_n[C]$, is defined as

$$\gamma_n[C] := \int_0^T dt\dot{\mathbf{R}}(t)\cdot\mathrm{i}\,\langle n, \mathbf{R}(t)|\nabla_R|n, \mathbf{R}(t)\rangle = \oint_C d\mathbf{R}\cdot\mathrm{i}\,\langle n, \mathbf{R}|\nabla_R|n, \mathbf{R}\rangle\,. \tag{3.10}$$

Defining the *Berry connection*

$$\mathbf{A}_n(\mathbf{R}) := -\mathrm{i}\,\langle n, \mathbf{R}|\nabla_R|n, \mathbf{R}\rangle\,, \tag{3.11}$$

and the *Berry curvature*

$$\mathbf{B}_n(\mathbf{R}) := \nabla_R \times \mathbf{A}_n(\mathbf{R}), \tag{3.12}$$

we can rewrite the Berry phase as

$$\gamma_n[C] = -\oint_C d\mathbf{R} \bullet \mathbf{A}_n(\mathbf{R}) = -\int_S d\mathbf{S} \bullet \mathbf{B}_n(\mathbf{R}), \tag{3.13}$$

where the last equality comes from Stokes' theorem.

We see that the Berry phase describes the accumulated phase factor of a quantum mechanical system after it completes a closed loop in parameter space.

We note that the Berry connection is a connection in the mathematical sense described in Section B.2. We follow the description in [16, sec 10.6.2]. Namely, let M be a manifold that describes the parameter space, and let $\mathbf{R} = (R_1, \ldots, R_k)$ be the local coordinate. At each point $\mathbf{R}$ of M we consider the normalized nth eigenstate of the Hamiltonian $H(\mathbf{R})$. Each such state is, as we know, represented by an equivalence class of states

$$[|\mathbf{R}\rangle] = \{g\,|\mathbf{R}\rangle : g \in U(1)\}\,. \tag{3.14}$$

At each point $\mathbf{R}$ of M, we have a $U(1)$-principal bundle $P(M, U(1))$ over the parameter space M. The projection is given by $p(g\,|\mathbf{R}\rangle) = \mathbf{R}$.

We can choose a section of $P(M, U(1))$ by fixing the phase of $|\mathbf{R}\rangle$ at each point $\mathbf{R} \in M$. Now, let $\sigma(\mathbf{R}) = |\mathbf{R}\rangle$ be a local section over a chart U of M. The canonical local trivialization is given by

$$\phi^{-1}(|\mathbf{R}\rangle) = (\mathbf{R}, e), \tag{3.15}$$

with e the unit element of $U(1)$. The right action of $U(1)$ gives us

$$\phi^{-1}(|\mathbf{R}\rangle \cdot g) = (\mathbf{R}, e)g = (\mathbf{R}, g). \tag{3.16}$$

So now we have defined the bundle structure, and we can move on to show why the Berry connection is a real connection. In a slightly more general notation, we let the Berry connection be given by

$$\mathcal{A} = \mathcal{A}_\mu dR^\mu, \tag{3.17}$$

where $\mathrm{d} = \frac{\partial}{\partial R^\mu} dR^\mu$ is the exterior derivative in $\mathbf{R}$-space.

Now, let U_i and U_j be overlapping charts of M and let $\sigma_i(\mathbf{R}) = |\mathbf{R}\rangle_i$ and $\sigma_j(\mathbf{R}) = |\mathbf{R}\rangle_j$ be the respective local sections. They are related by the transition function as $|\mathbf{R}\rangle_j = |\mathbf{R}\rangle_i\, t_{ij}(\mathbf{R})$. One can show that

$$\mathcal{A}_j(\mathbf{R}) = \mathcal{A}_i(\mathbf{R}) + t_{ij}(\mathbf{R})^{-1}\mathrm{d}t_{ij}(\mathbf{R}). \tag{3.18}$$

The set of one-forms $\{A_i\}$ with this transformation property, can be shown to define a connection on the principal bundle $P(M, U(1))$.

3.1.4 The TKNN-invariant

Now, the quantum Hall system is an example of topological insulator, and thus we are interested in finding ways to describe different phases that exist in this system. To do this, we use a topological invariant, called the TKNN-invariant (where TKNN stands for Thouless, Kohmoto, Nightingale

and den Nijs), to describe this system. We will now derive this invariant by calculating the Hall conductivity. To do this, we follow the approach in [5].

Consider a two-dimensional system of size $L \times L$ and let the system be in perpendicular electric and magnetic fields, where the electric field E is applied along the y-axis and the magnetic field B is applied along the z-axis.

Now, denote by H_0 the Hamiltonian of the system without the electric field, and let

$$H = H_0 - eEy, \tag{3.19}$$

where $V = -eEy$ naturally is the potential created by the electric field. We treat V as a perturbation of H_0, and use perturbation theory to approximate the eigenstate $|n\rangle_E$ as

$$|n\rangle_E = |n\rangle + \sum_{m \neq n} \frac{\langle m|(-eEy)|n\rangle}{E_n - E_m} |m\rangle + \dots. \tag{3.20}$$

Now we want to use this to approximate the current density along the x-axis. We have

$$\begin{aligned} \langle j_x \rangle_E &= \sum_n f(E_n)_E \langle n| \frac{ev_x}{L^2} |n\rangle_E \\ &= \langle j_x \rangle_{E=0} + \frac{1}{L^2} \sum_n f(E_n) \sum_{m \neq n} \Bigg(\frac{\langle n|ev_x|m\rangle \langle m|(-eEy)|n\rangle}{E_n - E_m} \\ &\qquad + \frac{\langle n|(-eEy)|m\rangle \langle m|ev_x|n\rangle}{E_n - E_m} \Bigg), \end{aligned} \tag{3.21}$$

where v_x is the electron velocity in the x-direction and $f(E_n)$ is the Fermi distribution function.

Now we have the Heisenberg equation of motion, described in equation (2.17), which states that

$$\frac{dy}{dt} = v_y = \frac{1}{\mathrm{i}\hbar} [y, H]. \tag{3.22}$$

Using this we get

$$\begin{aligned} \langle m|v_y|n\rangle &= \langle m| \frac{1}{\mathrm{i}\hbar} [y, H] |n\rangle = \frac{1}{\mathrm{i}\hbar} [\langle m|yH|n\rangle - \langle m|Hy|n\rangle] \\ &= \frac{1}{\mathrm{i}\hbar} [E_n \langle m|y|n\rangle - E_m \langle m|y|n\rangle] = \frac{1}{\mathrm{i}\hbar} (E_n - E_m) \langle m|y|n\rangle . \end{aligned} \tag{3.23}$$

This is equivalent to

$$\langle m|y|n\rangle = \frac{\mathrm{i}\hbar}{E_n - E_m} \langle m|v_y|n\rangle . \tag{3.24}$$

Using this together with equation (3.21), we get

$$\begin{aligned}
\sigma_{xy} &= \frac{\langle j_x \rangle_E - \langle j_x \rangle_{E=0}}{E} \\
&= \frac{1}{E}\frac{1}{L^2}\sum_n f(E_n) \sum_{m\neq n} \left(\frac{-eE \langle n|ev_x|m\rangle \frac{\mathrm{i}\hbar}{E_n - E_m} \langle m|v_y|n\rangle}{E_n - E_m} \right. \\
&\qquad \left. - \frac{eE\frac{\mathrm{i}\hbar}{E_m - E_n} \langle n|v_y|m\rangle \langle m|ev_x|n\rangle}{E_n - E_m} \right) \\
&= \frac{1}{EL^2}\sum_n f(E_n) \sum_{m\neq n} \left(\frac{-e^2\mathrm{i}\hbar E \langle n|v_x|m\rangle \langle m|v_y|n\rangle + e^2 E\mathrm{i}\hbar \langle n|v_y|m\rangle \langle m|v_x|n\rangle}{(E_n - E_m)^2} \right) \\
&= -\frac{\mathrm{i}\hbar e^2}{L^2}\sum_n \sum_{m\neq n} f(E_n) \frac{\langle n|v_x|m\rangle \langle m|v_y|n\rangle - \langle n|v_y|m\rangle \langle m|v_x|n\rangle}{(E_n - E_m)^2}.
\end{aligned} \tag{3.25}$$

The systems we are considering are crystals, and thus we have a periodic potential. This means that we can rewrite everything in terms of Bloch functions, i.e. let

$$|n\rangle = \sum_{\mathbf{k}} \exp(\mathrm{i}\mathbf{k} \bullet \mathbf{r}) \, |u_{n\mathbf{k}}\rangle \, . \tag{3.26}$$

Noting that the exponentials will all cancel out when inserting the Bloch functions into equation (3.25) we have

$$\sigma_{xy} = -\frac{\mathrm{i}\hbar e^2}{L^2}\sum_n \sum_{m\neq n} \sum_{\mathbf{k}} f(E_{n\mathbf{k}}) \frac{\langle u_{n\mathbf{k}}|v_x|u_{m\mathbf{k}}\rangle \langle u_{m\mathbf{k}}|v_y|u_{n\mathbf{k}}\rangle - \langle u_{n\mathbf{k}}|v_y|u_{m\mathbf{k}}\rangle \langle u_{m\mathbf{k}}|v_x|u_{n\mathbf{k}}\rangle}{(E_{n\mathbf{k}} - E_{m\mathbf{k}})^2}. \tag{3.27}$$

According to [5], we have

$$\langle u_{m\mathbf{k}'}|v_\mu|u_{n\mathbf{k}}\rangle = \frac{1}{\hbar}(E_{n\mathbf{k}} - E_{m\mathbf{k}'}) \langle u_{m\mathbf{k}'}|\frac{\partial}{\partial k_\mu}|u_{n\mathbf{k}}\rangle \, . \tag{3.28}$$

Inserting this into equation (3.27), we get

$$\begin{aligned}
\sigma_{xy} = -\frac{\mathrm{i}e^2}{\hbar L^2}\sum_{\mathbf{k}} \sum_n \sum_{m\neq n} f(E_{n\mathbf{k}}) \Big(&\langle u_{n\mathbf{k}}|\frac{\partial}{\partial k_x}|u_{m\mathbf{k}}\rangle \langle u_{m\mathbf{k}}|\frac{\partial}{\partial k_y}|u_{n\mathbf{k}}\rangle \\
&- \langle u_{n\mathbf{k}}|\frac{\partial}{\partial k_y}|u_{m\mathbf{k}}\rangle \langle u_{m\mathbf{k}}|\frac{\partial}{\partial k_x}|u_{n\mathbf{k}}\rangle \Big) \, .
\end{aligned} \tag{3.29}$$

This gives us

$$\sigma_{xy} = -\frac{\mathrm{i}e^2}{\hbar L^2}\sum_{\mathbf{k}} \sum_{n\neq m} f(E_{n\mathbf{k}}) \left(\frac{\partial}{\partial k_x} \langle u_{n\mathbf{k}}|\frac{\partial}{\partial k_y}u_{n\mathbf{k}}\rangle - \frac{\partial}{\partial k_y} \langle u_{n\mathbf{k}}|\frac{\partial}{\partial k_x}u_{n\mathbf{k}}\rangle \right) . \tag{3.30}$$

For Bloch states, the Berry connection is given by

$$\mathbf{a}_n(\mathbf{k}) = -\mathrm{i}\,\langle u_{n\mathbf{k}}|\nabla_{\mathbf{k}}|u_{n\mathbf{k}}\rangle = -\mathrm{i}\,\langle u_{n\mathbf{k}}|\frac{\partial}{\partial\mathbf{k}}|u_{n\mathbf{k}}\rangle\,. \tag{3.31}$$

This means we can express the Hall conductivity in the following way:

$$\sigma_{xy} = \nu\frac{e^2}{h}, \tag{3.32}$$

where ν is given by

$$\nu = \sum_n \int_{BZ} \frac{d^2\mathbf{k}}{2\pi}\left(\frac{\partial a_{n,y}}{\partial k_x} - \frac{\partial a_{n,x}}{\partial k_y}\right). \tag{3.33}$$

We express ν as

$$\nu = \sum_n \nu_n, \tag{3.34}$$

where ν_n is the contribution from the nth band. Now one can show that it is related to the Berry phase, defined in equation (3.13), by

$$\nu_n = \int_{BZ} \frac{d^2\mathbf{k}}{2\pi}\left(\frac{\partial a_{n,y}}{\partial k_x} - \frac{\partial a_{n,x}}{\partial k_y}\right) = \frac{1}{2\pi}\oint_{\partial BZ} d\mathbf{k}\bullet\mathbf{a}_n(\mathbf{k}) = -\frac{1}{2\pi}\gamma_n\left[\partial BZ\right]. \tag{3.35}$$

The change in phase of the wave function after encircling the Brillouin zone boundary must be an integer multiple of 2π. This means that

$$\gamma_n\left[\partial BZ\right] = 2\pi m, \tag{3.36}$$

where m is an integer. Thus ν_n must be an integer, and thus σ_{xy} is quantized to integer multiples of e^2/h. The integer ν is called the *TKNN-invariant* and is the topological invariant we use to differ between the different phases in the integer quantum Hall system.

3.2 Time-reversal symmetry in topological insulators

So far, we have been solely focused on the quantum Hall system. We will now study more general systems that are invariant under time-reversal symmetry.

3.2.1 Time-reversal symmetry and the Bloch Hamiltonian

Let $\mathcal{H}$ be the total Hamiltonian of a periodic spin-1/2 system, i.e.

$$\mathcal{H}\,|\psi_{n\mathbf{k}}\rangle = E_{n\mathbf{k}}\,|\psi_{n\mathbf{k}}\rangle\,. \tag{3.37}$$

According to Bloch's theorem we can rewrite $|\psi_{n\mathbf{k}}\rangle$ as

$$|\psi_{n\mathbf{k}}\rangle = e^{-\mathrm{i}\mathbf{k}\bullet\mathbf{r}}\,|u_{n\mathbf{k}}\rangle\,, \tag{3.38}$$

where $|u_{n\mathbf{k}}\rangle$ is the eigenstate of the Bloch Hamiltonian

$$H(\mathbf{k}) = e^{-\mathrm{i}\mathbf{k}\bullet\mathbf{r}}\mathcal{H}e^{\mathrm{i}\mathbf{k}\bullet\mathbf{r}}. \tag{3.39}$$

The eigenstate $|u_{n\mathbf{k}}\rangle$ satisfies the reduced Schrödinger equation

$$H(\mathbf{k})\,|u_{n\mathbf{k}}\rangle = E_{n\mathbf{k}}\,|u_{n\mathbf{k}}\rangle\,. \tag{3.40}$$

Since we are dealing with half-integer spin, we have

$$\Theta^2 = -1, \tag{3.41}$$

as is described in Section 2.3.

When $\mathcal{H}$ preserves time reversal symmetry we have

$$[\mathcal{H}, \Theta] = 0. \tag{3.42}$$

This means that

$$\Theta\mathcal{H}\Theta^{-1} = \mathcal{H}. \tag{3.43}$$

One can argue that

$$\Theta\exp(\mathrm{i}\mathbf{k}\bullet\mathbf{r}) = \exp(-\mathrm{i}\mathbf{k}\bullet\mathbf{r})\Theta, \tag{3.44}$$

which gives us

$$\Theta H(\mathbf{k})\Theta^{-1} = \Theta e^{-\mathrm{i}\mathbf{k}\bullet\mathbf{r}}\mathcal{H}e^{\mathrm{i}\mathbf{k}\bullet\mathbf{r}}\Theta^{-1} = e^{\mathrm{i}\mathbf{k}\bullet\mathbf{r}}\Theta\mathcal{H}\Theta e^{-\mathrm{i}\mathbf{k}\bullet\mathbf{r}} = e^{\mathrm{i}\mathbf{k}\bullet\mathbf{r}}\mathcal{H}e^{-\mathrm{i}\mathbf{k}\bullet\mathbf{r}} = H(-\mathbf{k}). \tag{3.45}$$

This result implies that at momenta $\mathbf{k}$ that satisfy $H(\mathbf{k}) = H(-\mathbf{k})$, the system is time-reversal invariant. Such points are called *time-reversal invariant momenta*, and exist because of the periodicity of the Brillouin zone.

Now, let $\psi_n(\mathbf{k})$ be an eigenstate of $\mathcal{H}$, i.e. let

$$\mathcal{H}\psi_n(\mathbf{k}) = E\psi_n(\mathbf{k}), \tag{3.46}$$

for some E. Now consider the action of $\mathcal{H}$ on the time-reversed state $\Theta\psi$. We have

$$\mathcal{H}\Theta\psi_n(\mathbf{k}) = \Theta\mathcal{H}\psi_n(\mathbf{k}) = \Theta E\psi_n(\mathbf{k}) = E\Theta\psi_n(\mathbf{k}). \tag{3.47}$$

This result means that if $\psi_n(\mathbf{k})$ is an eigenstate of $\mathcal{H}$, then this is true also for $\Theta\psi_n(\mathbf{k})$.

Now we want to show that these two states are different so that we always have degeneracy in these systems. Namely, assume that they are the same state, i.e. that

$$\Theta\,|n\rangle = e^{\mathrm{i}\alpha}\,|n\rangle\,, \tag{3.48}$$

for some $\alpha \in \mathbb{R}$. Applying Θ twice then gives us

$$\Theta^2\,|n\rangle = \Theta(e^{\mathrm{i}\alpha}\,|n\rangle) = e^{-\mathrm{i}\alpha}e^{\mathrm{i}\alpha}\,|n\rangle = |n\rangle\,. \tag{3.49}$$

This result implies that

$$\Theta^2 = 1, \tag{3.50}$$

which is a contradiction in the case of spin-1/2-systems that we are dealing with. Thus the states must be different, and thus the energy bands of a time-reversal symmetric system come in pairs. These pairs are called *Kramers pairs*. These pairs are degenerate at time-reversal invariant momenta, described in equation (3.45)

A suitable matrix representation of the time-reversal operator is

$$w_{\alpha\beta}(\mathbf{k}) = \langle u_{\alpha,-\mathbf{k}}|\Theta|u_{\beta,\mathbf{k}}\rangle\,, \tag{3.51}$$

where α and β are band indices. This matrix relates the Bloch states $|u_{\alpha,-\mathbf{k}}\rangle$ and $|u_{\beta,\mathbf{k}}\rangle$ via

$$|u_{\alpha,-\mathbf{k}}\rangle = \sum_\beta w^*_{\alpha\beta}(\mathbf{k})\Theta\,|u_{\beta,\mathbf{k}}\rangle\,. \tag{3.52}$$

In [5], it is claimed that $w_{\alpha\beta}(\mathbf{k})$ is a unitary matrix. This can be seen in the following way:

$$\begin{aligned}\sum_\alpha w^\dagger_{\gamma\alpha}(\mathbf{k})w_{\alpha\beta}(\mathbf{k}) &= \sum_\alpha \langle u_\alpha(-\mathbf{k})|\Theta|u_\gamma(\mathbf{k})\rangle^* \langle u_\alpha(-\mathbf{k})|\Theta|u_{\beta(\mathbf{k})}\rangle \\ &= \sum_\alpha \langle \Theta u_\gamma(\mathbf{k})|u_\alpha(-\mathbf{k})\rangle \langle u_\alpha(-\mathbf{k})|\Theta|u_\beta(\mathbf{k})\rangle \\ &= \langle \Theta u_\gamma(\mathbf{k})|\Theta|u_\beta(\mathbf{k})\rangle = \langle u_\beta(\mathbf{k})|u_\gamma(\mathbf{k})\rangle = \delta_{\beta\gamma}.\end{aligned} \tag{3.53}$$

Using equation (2.25), we also show that the following property stated in [5] is true:

$$w_{\beta\alpha}(-\mathbf{k}) = \langle u_\beta(\mathbf{k})|\Theta|u_\alpha(-\mathbf{k})\rangle = -\langle u_\alpha(-\mathbf{k})|\Theta|u_\beta(\mathbf{k})\rangle = -w_{\alpha\beta}(\mathbf{k}). \tag{3.54}$$

This last equation implies that w is an antisymmetric matrix at time-reversal invariant momenta $\mathbf{k} = \mathbf{\Lambda_i}$. I.e. we have

$$w_{\beta\alpha}(\mathbf{\Lambda}_i) = -w_{\alpha\beta}(\mathbf{\Lambda}_i). \tag{3.55}$$

We are not only interested in the w-matrix. Another important matrix is the $U(2)$ Berry connection matrix (which in reality is a collection of three matrices) defined in the following way:

$$\mathbf{a}_{\alpha\beta}(\mathbf{k}) := -\mathrm{i}\,\langle u_{\alpha,\mathbf{k}}|\nabla_\mathbf{k}|u_{\beta,\mathbf{k}}\rangle\,. \tag{3.56}$$

In [5] it is claimed that $\mathbf{a}^*_{\beta\alpha}(\mathbf{k}) = \mathbf{a}_{\alpha\beta}(\mathbf{k})$. This can be realized in the following way:

$$\begin{aligned}\mathbf{a}^*_{\beta\alpha}(\mathbf{k}) &= (-\mathrm{i}\,\langle u_\beta(\mathbf{k})|\nabla_\mathbf{k}|u_\alpha(\mathbf{k})\rangle)^* = \mathrm{i}\,\langle \nabla_\mathbf{k} u_\alpha(\mathbf{k})|u_\beta(\mathbf{k})\rangle \\ &= -\mathrm{i}\,\langle u_\alpha(\mathbf{k})|\nabla_\mathbf{k}|u_\beta(\mathbf{k})\rangle = \mathbf{a}_{\alpha\beta}(\mathbf{k}),\end{aligned} \tag{3.57}$$

We also have

$$\begin{aligned}\mathbf{a}_{\alpha\beta}(-\mathbf{k}) &= -\mathrm{i}\,\langle u_\alpha(-\mathbf{k})|\nabla_{-\mathbf{k}}|u_\beta(-\mathbf{k})\rangle = \mathrm{i}\,\langle u_\alpha(-\mathbf{k})|\nabla_\mathbf{k}|u_\beta(-\mathbf{k})\rangle \\ &= \mathrm{i}\,\langle \sum_\gamma w^*_{\alpha\gamma}(\mathbf{k})\Theta u_\gamma(\mathbf{k})|\nabla_\mathbf{k}|\sum_\mu w^*_{\beta\mu}(\mathbf{k})\Theta u_\mu(\mathbf{k})\rangle \\ &= \mathrm{i}\,\langle \sum_\gamma w^*_{\alpha\gamma}(\mathbf{k})\Theta u_\gamma(\mathbf{k})|\sum_\mu \left[\nabla_\mathbf{k}(w^*_{\beta\mu}(\mathbf{k}))\Theta u_\mu(\mathbf{k}) + w^*_{\beta\mu}(\mathbf{k})\nabla_\mathbf{k}(\Theta u_\mu(\mathbf{k}))\right]\rangle \\ &= \mathrm{i}\sum_{\gamma\mu} w_{\alpha\gamma}(\mathbf{k})\nabla_\mathbf{k}(w^*_{\beta\mu}(\mathbf{k}))\,\langle \Theta u_\gamma(\mathbf{k})|\Theta u_\mu(\mathbf{k})\rangle + \mathrm{i}\sum_{\gamma\mu} w_{\alpha\gamma}(\mathbf{k})w^*_{\beta\mu}(\mathbf{k})\,\langle \Theta u_\gamma(\mathbf{k})|\nabla_k|\Theta u_\mu(\mathbf{k})\rangle\end{aligned} \tag{3.58}$$

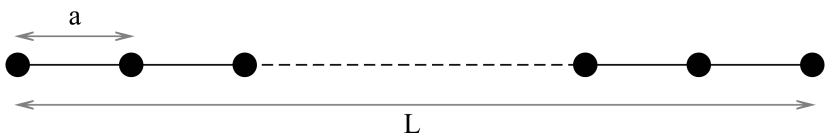

Figure 2: The one-dimensional system for which the topological invariant is derived.

Now, in [17] and [18, eq. (10.40)] it is implied that

$$\langle \Theta u_\gamma(\mathbf{k}) | \nabla_k | \Theta u_\mu(\mathbf{k}) \rangle = - \langle u_\mu(\mathbf{k}) | \nabla_k | u_\gamma(\mathbf{k}) \rangle . \tag{3.59}$$

This means that

$$\begin{aligned}
\mathbf{a}_{\alpha\beta}(-\mathbf{k}) &= \mathrm{i} \sum_{\gamma\mu} w_{\alpha\gamma}(\mathbf{k}) \nabla_{\mathbf{k}} (w^*_{\beta\mu}(\mathbf{k})) \langle u_\mu(\mathbf{k}) | u_\gamma(\mathbf{k}) \rangle - \mathrm{i} \sum_{\gamma\mu} w_{\alpha\gamma}(\mathbf{k}) w^*_{\beta\mu}(\mathbf{k}) \langle u_\mu(\mathbf{k}) | \nabla_k | u_\gamma(\mathbf{k}) \rangle \\
&= \mathrm{i} \sum_{\gamma} w_{\alpha\gamma}(\mathbf{k}) \nabla_{\mathbf{k}} w^*_{\beta\gamma}(\mathbf{k}) + \sum_{\gamma\mu} w_{\alpha\gamma}(\mathbf{k}) w^*_{\beta\mu}(\mathbf{k}) \mathbf{a}_{\mu\gamma}(\mathbf{k}) \\
&= \mathrm{i} \sum_{\gamma} w_{\alpha\gamma}(\mathbf{k}) \nabla_{\mathbf{k}} w^*_{\beta\gamma}(\mathbf{k}) + \sum_{\gamma\mu} w_{\alpha\gamma}(\mathbf{k}) \mathbf{a}^*_{\gamma\mu}(\mathbf{k}) w^*_{\beta\mu}(\mathbf{k}) \\
&= \mathrm{i} \sum_{\gamma} w_{\alpha\gamma}(\mathbf{k}) \nabla_{\mathbf{k}} w^\dagger_{\gamma\beta}(\mathbf{k}) + \sum_{\gamma\mu} w_{\alpha\gamma}(\mathbf{k}) \mathbf{a}^*_{\gamma\mu}(\mathbf{k}) w^\dagger_{\mu\beta}(\mathbf{k}),
\end{aligned} \tag{3.60}$$

as is stated in [5]. Rewriting this in terms of matrices, we get

$$\mathbf{a}(-\mathbf{k}) = w(\mathbf{k}) \mathbf{a}^*(\mathbf{k}) w^\dagger(\mathbf{k}) + \mathrm{i} w(\mathbf{k}) \nabla_{\mathbf{k}} w^\dagger(\mathbf{k}). \tag{3.61}$$

Taking the trace of this equation, we get

$$\mathrm{tr}\left[\mathbf{a}(-\mathbf{k})\right] = \mathrm{tr}\left[\mathbf{a}^*(\mathbf{k})\right] + \mathrm{tr}\left[w(\mathbf{k}) \nabla_{\mathbf{k}} w^\dagger(\mathbf{k}) \right]. \tag{3.62}$$

Now, we have that $\mathbf{a}_{\alpha\beta} = \mathbf{a}^*_{\beta\alpha}$, so

$$\mathrm{tr}\left[\mathbf{a}\right] = \mathrm{tr}\left[\mathbf{a}^*\right]. \tag{3.63}$$

Also, since $w w^\dagger = I$, we have

$$w \nabla w^\dagger = -(\nabla w) w^\dagger. \tag{3.64}$$

Using this and making the transformation $\mathbf{k} \to -\mathbf{k}$ in equation (3.62), we get

$$\mathrm{tr}\left[\mathbf{a}(\mathbf{k})\right] = \mathrm{tr}\left[\mathbf{a}(-\mathbf{k})\right] + \mathrm{tr}\left[w^\dagger(\mathbf{k}) \nabla_{\mathbf{k}} w(\mathbf{k}) \right]. \tag{3.65}$$

3.2.2 $\mathbb{Z}_2$ time-reversal polarization

In this section we derive the topological invariant for electron systems that preserve time-reversal symmetry. We do this as in [5] and begin by describing a one-dimensional crystalline system with length L and lattice constant $a = 1$, see Figure 2.

To begin with, we only consider two energy bands that form a Kramers pair, and denote these states by $|u_1(k)\rangle$ and $|u_2(k)\rangle$. Assume that the band parameters change adiabatically with time

and that they return to their original value at time $t = T$. Assume further that the Hamiltonian of the system satisfies the following:

$$H\left[t+T\right] = H\left[t\right], \tag{3.66}$$

$$H\left[-t\right] = \Theta H\left[t\right]\Theta^{-1}. \tag{3.67}$$

The electronic polarization $\mathbf{P}$ for a three-dimensional crystalline material with N occupied bands is given by [19]

$$\mathbf{P} = -\frac{2\mathrm{i}}{(2\pi)^3}\int_{BZ} d\mathbf{k}\sum_{n=1}^{N}\left\langle u_{n\mathbf{k}}|\nabla_{\mathbf{k}}u_{n\mathbf{k}}\right\rangle. \tag{3.68}$$

In our case, this reduces to [5]

$$P_\rho = \int_{-\pi}^{\pi}\frac{dk}{2\pi}A(k), \tag{3.69}$$

where

$$A(k) = -\mathrm{i}\left\langle u_1(k)|\nabla_k|u_1(k)\right\rangle - \mathrm{i}\left\langle u_2(k)|\nabla_k|u_2(k)\right\rangle = a_{11}(k) + a_{22}(k) = \mathrm{tr}\left[\mathbf{a}(k)\right]. \tag{3.70}$$

Note that it makes sense to talk about polarization in this case because the material is an insulator. [19]

We continue to follow [5] and note that we can divide the charge polarization P_ρ into two parts, one from each band. We call such a part a partial polarization, and define it by

$$P_i = \int_{-\pi}^{\pi}\frac{dk}{2\pi}a_{ii}(k), \tag{3.71}$$

which gives us

$$P_\rho = P_1 + P_2. \tag{3.72}$$

Now, one can also define the time-reversal polarization [17]

$$P_\theta := P_1 - P_2 = 2P_1 - P_\rho. \tag{3.73}$$

This can be interpreted as the difference in charge polarization between spin up and spin down bands, since $|u_1(k)\rangle$ and $|u_2(k)\rangle$ form a Kramers pair.

The properties of the Hamiltonian given in equations (3.66) and (3.67) result in the system being time-reversal symmetric at times $t = 0$ and $t = T/2$. At these times, the Kramers degeneracy must be observed at all values of k, which means that

$$\Theta\left|u_2(k)\right\rangle = e^{-\mathrm{i}\chi(k)}\left|u_1(-k)\right\rangle, \tag{3.74}$$

and

$$\Theta\left|u_1(k)\right\rangle = -e^{-\mathrm{i}\chi(k)}\left|u_2(-k)\right\rangle. \tag{3.75}$$

This gives us the following w-matrix:

$$w(k) = \begin{pmatrix} 0 & e^{-\mathrm{i}\chi(k)} \\ -e^{-\mathrm{i}\chi(k)} & 0 \end{pmatrix}. \tag{3.76}$$

Now we want to calculate P_1 at the time-reversal symmetric times. We know that

$$a_{ii}(k) = -\mathrm{i}\,\langle u_i(k)|\nabla_k|u_i(k)\rangle \tag{3.77}$$

This means that

$$a_{ii}(-k) = -\mathrm{i}\,\langle u_i(-k)|\nabla_{-k}|u_i(-k)\rangle = \mathrm{i}\,\langle u_i(-k)|\nabla_k|u_i(-k)\rangle\,. \tag{3.78}$$

From equation (3.74), we then get

$$\begin{aligned} a_{11}(-k) &= \mathrm{i}\,\langle e^{\mathrm{i}\chi(k)}\Theta u_2(k)|\nabla_k|e^{\mathrm{i}\chi(k)}\Theta u_2(k)\rangle = \mathrm{i}e^{-\mathrm{i}\chi(k)}\,\langle \Theta u_2(k)|\nabla_k(e^{\mathrm{i}\chi(k)}\Theta u_2(k))\rangle \\ &= \mathrm{i}e^{-\mathrm{i}\chi(k)}\,\langle \Theta u_2(k)|\nabla_k(e^{\mathrm{i}\chi(k)})\Theta u_2(k)\rangle + \mathrm{i}e^{-\mathrm{i}\chi(k)}\,\langle \Theta u_2(k)|e^{\mathrm{i}\chi(k)}\nabla_k(\Theta u_2(k))\rangle \\ &= -\nabla_k(\chi(k))\,\langle \Theta u_2(k)|\Theta u_2(k)\rangle + \mathrm{i}\,\langle \Theta u_2(k)|\nabla_k(\Theta u_2(k))\rangle \\ &= -\nabla_k(\chi(k)) - \mathrm{i}\,\langle \nabla_k(u_2(k))|u_2(k)\rangle = -\nabla_k(\chi(k)) - \mathrm{i}\,\langle u_2(k)|\nabla_k|u_2(k)\rangle \\ &= -\nabla_k(\chi(k)) + a_{22}(k), \end{aligned} \tag{3.79}$$

where we have used the result

$$\langle \Theta u_2(k)|\Theta\nabla_k(u_2(k))\rangle = \langle u_2(k)|\nabla_k|u_2(k)\rangle\,, \tag{3.80}$$

stated e.g. in [17] and [18, eq. 10.40].

Therefore, since we are in one dimension, we have

$$a_{11}(-k) = a_{22}(k) - \frac{\partial}{\partial k}\chi(k), \tag{3.81}$$

which using equation (3.71) gives us

$$\begin{aligned} P_1 &= \frac{1}{2\pi}\left(\int_0^\pi dk\, a_{11} + \int_{-\pi}^0 dk\, a_{11}\right) = \frac{1}{2\pi}\int_0^\pi dk\,\left(a_{11} + a_{22} - \frac{\partial}{\partial k}\chi(k)\right) \\ &= \int_0^\pi \frac{dk}{2\pi}A(k) - \frac{1}{2\pi}\left[\chi(\pi) - \chi(0)\right]. \end{aligned} \tag{3.82}$$

Rewriting $\chi(k)$ in terms of a w-matrix element, we have from equation (3.76) that

$$\chi(k) = \mathrm{i}\log w_{12}(k), \tag{3.83}$$

which gives us

$$P_1 = \int_0^\pi \frac{dk}{2\pi}A(k) - \frac{\mathrm{i}}{2\pi}\log\left(\frac{w_{12}(\pi)}{w_{12}(0)}\right). \tag{3.84}$$

This in turn, using equation (3.69) means that

$$P_\theta = 2P_1 - P_\rho = \int_0^\pi \frac{dk}{2\pi}\left[A(k) - A(-k)\right] - \frac{\mathrm{i}}{\pi}\log\left(\frac{w_{12}(\pi)}{w_{12}(0)}\right). \tag{3.85}$$

From equation (3.65), we have an expression for $\mathrm{tr}\,[\mathbf{a}(\mathbf{k})]$. Using $A(k) = \mathrm{tr}\,[\mathbf{a}(\mathbf{k})]$, and inserting this into (3.85), we get

$$\begin{aligned} P_\theta &= \int_0^\pi \frac{dk}{2\pi} \mathrm{i}\,\mathrm{tr}\left[w^\dagger(k)\frac{\partial}{\partial k}w(k)\right] - \frac{\mathrm{i}}{\pi}\log\left(\frac{w_{12}(\pi)}{w_{12}(0)}\right) \\ &= \mathrm{i}\int_0^\pi \frac{dk}{2\pi}\frac{\partial}{\partial k}\log(\det\,[w(k)]) - \frac{\mathrm{i}}{\pi}\log\left(\frac{w_{12}(\pi)}{w_{12}(0)}\right) \\ &= \frac{\mathrm{i}}{\pi}\frac{1}{2}\log\left(\frac{\det\,[w(\pi)]}{\det\,[w(0)]}\right) - \frac{\mathrm{i}}{\pi}\log\left(\frac{w_{12}(\pi)}{w_{12}(0)}\right). \end{aligned} \tag{3.86}$$

Now, in this case we have $\det\,[w(k)] = w_{12}(k)^2$, so

$$P_\theta = \frac{1}{\mathrm{i}\pi}\log\left(\frac{\sqrt{w_{12}(0)^2}}{w_{12}(0)}\cdot\frac{w_{12}(\pi)}{\sqrt{w_{12}(\pi)^2}}\right). \tag{3.87}$$

Since the square root is always positive, it is clear that the argument of the logarithm is either $+1$ or -1. This means that

$$P_\theta = \begin{cases} 0 \bmod 2, & \text{if } \mathrm{sgn}(w_{12}(\pi)) = \mathrm{sgn}(w_{12}(0)) \\ 1 \bmod 2, & \text{if } \mathrm{sgn}(w_{12}(\pi)) = -\mathrm{sgn}(w_{12}(0)). \end{cases} \tag{3.88}$$

Physically, the two different values of P_θ correspond to the two different polarization states that the system can take on at times $t = 0$ and $t = T/2$.

It is interesting to consider the change of P_θ at intermediate times. (And we note that the system is not time-reversal invariant between the times.) We let

$$\Delta = (P_\theta(T/2) - P_\theta(0)) \bmod 2. \tag{3.89}$$

Equation (3.87) gives us the following way to express Δ:

$$(-1)^\Delta = \prod_{i=1}^{4}\frac{w_{12}(\Lambda_i)}{\sqrt{w_{12}(\Lambda_1)^2}}, \tag{3.90}$$

where $\Lambda_1 = (0,0)$, $\Lambda_2 = (\pi,0)$, $\Lambda_3 = (0,T/2)$ and $\Lambda_4 = (\pi,T/2)$.

It follows from the construction that Δ is a topological invariant, which can take the values 0 and 1. The interesting question is now what physically distinguishes the phases for which $\Delta = 0$ and $\Delta = 1$. This is discussed in [17]. The two-dimensional phase space (k,t) forms a torus because of the periodic boundary conditions, and the topological invariant Δ characterizes the mapping from this torus to the space of wave functions. In [17] it is argued that when $\Delta = 1$, the system behaves like a spin pump which pumps spin from one end of the system to the other. This connects back to the fact that P_θ is interpreted as the difference in charge polarization between spin up and spin down bands, see the definition of P_θ in equation (3.73).

One example of this phenomenon is the quantum spin Hall effect placed on a cylinder. In this case the system describes a sort of adiabatic pump as a function of the magnetic flux through

the cylinder. When S_z is conserved, increasing the magnetic flux by one flux quantum, causes a transfer of spin from one end of the cylinder to the other. In all cases, for such a pump to exist, there must be a conservation of spin. Such a conservation of spin is introduced through time reversal symmetry, which gives us the result above [17].

Now, all of the above refers to a system with only two energy bands. We want to extend this to a more general case, and since we know that the bands come in Kramers pairs, we consider a system with $2N$ occupied bands that form N Kramers pairs. For each such pair n, we have the following relations between the wave functions at time-reversal symmetric times:

$$\Theta \left| u_2^n(k) \right\rangle = e^{-\mathrm{i}\chi_n(k)} \left| u_1^n(-k) \right\rangle , \tag{3.91}$$

and

$$\Theta \left| u_1^n(k) \right\rangle = -e^{-\mathrm{i}\chi_n(k)} \left| u_2^n(-k) \right\rangle . \tag{3.92}$$

This means that the w-matrix is a $2N \times 2N$ block diagonal matrix

$$w(k) = \begin{pmatrix} M_1 & & & \\ & M_2 & & \\ & & \ddots & \\ & & & M_N \end{pmatrix} , \tag{3.93}$$

where

$$M_i = \begin{pmatrix} 0 & e^{-\mathrm{i}\chi_i(k)} \\ -e^{-\mathrm{i}\chi_i(k)} & 0 \end{pmatrix} . \tag{3.94}$$

For simplicity, we now let $T = 2\pi$. This means that $w(t = 0)$ and $w(t = \pi)$ become anti-symmetric, which enables us to calculate the Pfaffian of the matrix (see Section B.1),

$$\mathrm{Pf}\left[w(\Lambda_i) \right] = w_{12}(\Lambda_i) w_{34}(\Lambda_i) \dots w_{2N-1,2N}(\Lambda_i) = \exp\left[-\mathrm{i} \sum_{n=1}^{N} \chi_n(\Lambda_i) \right] . \tag{3.95}$$

According to [5], it is straightforward to extend the two-band arguments to this present case, and here we show how this can be done. We get

$$P_\rho = \int_{-\pi}^{\pi} \frac{dk}{2\pi} A(k), \tag{3.96}$$

where $A(k)$ now is given by

$$A(k) = -\sum_n \left[\mathrm{i} \left\langle u_n^1(k) | \nabla_k | u_n^1(k) \right\rangle + \mathrm{i} \left\langle u_n^2(k) | \nabla_k | u_n^2(k) \right\rangle \right] . \tag{3.97}$$

Similarly, we write

$$P_\rho = \sum_n \left[P_n^1 + P_n^2 \right] , \tag{3.98}$$

where

$$P_n^i = \int_{-\pi}^{\pi} \frac{dk}{2\pi} a_n^i(k). \tag{3.99}$$

Here we have set

$$a_n^i(k) = -\mathrm{i}\,\langle u_n^i(k)|\nabla_k|u_n^i(k)\rangle\,. \tag{3.100}$$

By the same argument as for the case with only two bands, we get

$$a_n^1(-k) = a_n^2(k) - \frac{\partial}{\partial k}\chi_n(k). \tag{3.101}$$

This means we have

$$\begin{aligned} P_1 &= \frac{1}{2\pi}\int_0^\pi \sum_n \left(a_n^1 + a_n^2 - \frac{\partial}{\partial k}\chi_n(k)\right) \\ &= \int_0^\pi \frac{dk}{2\pi}A(k) - \frac{1}{2\pi}\sum_{n=1}^N \left[\chi_n(\pi) - \chi_n(0)\right] \\ &= \int_0^\pi \frac{dk}{2\pi}A(k) - \frac{\mathrm{i}}{2\pi}\log\left(\frac{\mathrm{Pf}\left[w(\pi)\right]}{\mathrm{Pf}\left[w(0)\right]}\right), \end{aligned} \tag{3.102}$$

where we have used the expression for the Pfaffian in equation (3.95) in the last step.

Using this, we also get the following time-reversal polarization:

$$P_\theta = \frac{1}{\mathrm{i}\pi}\log\left(\frac{\sqrt{\det\left[w(0)\right]}}{\mathrm{Pf}\left[w(0)\right]}\cdot\frac{\mathrm{Pf}\left[w(\pi)\right]}{\sqrt{\det\left[w(\pi)\right]}}\right), \tag{3.103}$$

and analogously as before we see that it is given by

$$P_\theta = \begin{cases} 0 \bmod 2, & \text{if } \mathrm{sgn}(\mathrm{Pf}\left[w(\pi)\right]) = \mathrm{sgn}(\mathrm{Pf}\left[w(0)\right]) \\ 1 \bmod 2, & \text{if } \mathrm{sgn}(\mathrm{Pf}\left[w(\pi)\right]) = -\mathrm{sgn}(\mathrm{Pf}\left[w(0)\right]). \end{cases} \tag{3.104}$$

We thus get that the $\mathbb{Z}_2$ topological invariant is given by

$$(-1)^\Delta = \prod_{i=1}^4 \frac{\mathrm{Pf}\left[w(\Lambda_i)\right]}{\sqrt{\det\left[w(\Lambda_i)\right]}}, \tag{3.105}$$

and as before we get a classification into two phases, where the one with $\Delta = 1$ can be interpreted as a spin pump.

3.2.3 Extension to three-dimensional systems

The previous discussion was limited to two-dimensional systems, and it is of interest to extend the theory to the three-dimensional case. According to [5] this can be done using a homotopy argument described in [20], but we will instead use the approach described in [5].

Suppose we have a cubic system with lattice constant $a = 1$. In the three-dimensional Brillouin zone of this system, we let the time-reversal invariant momenta be given by $\Lambda_{0,0,0}$, $\Lambda_{0,0,\pi}$, $\Lambda_{0,\pi,0}$, $\Lambda_{\pi,0,0}$, $\Lambda_{0,\pi,\pi}$, $\Lambda_{\pi,0,\pi}$, $\Lambda_{\pi,\pi,0}$ and $\Lambda_{\pi,\pi,\pi}$. At these points, as before the Hamiltonian becomes time-reversal symmetric and each Kramers pair of bands becomes degenerate.

The six planes $x = 0$, $x = \pi$, $y = 0$, $y = \pi$, $z = 0$, $z = \pi$ all have the same symmetries as the two-dimensional Brillouin zone, so we have altogether six $\mathbb{Z}_2$ invariants, one for each plane, which we denote by x_0, x_1, y_0, y_1, z_0 and z_1. These are not all independent of each other; we have

$$x_0 x_1 = y_0 y_1 = z_0 z_1. \tag{3.106}$$

These constraining relations show us that there are only four independent invariants in the system.

For each time-reversal invariant momentum, Λ_i, define

$$\delta(\Lambda_i) := \frac{\mathrm{Pf}\left[w(\Lambda_i)\right]}{\sqrt{\det\left[w(\Lambda_i)\right]}}. \tag{3.107}$$

Now the four $\mathbb{Z}_2$ invariants, we denote by ν_0, ν_1, ν_2 and ν_3, and they can be expressed as

$$(-1)^{\nu_0} = \prod_{n_j=0,\pi} \delta(\Lambda_{n_1,n_2,n_3}), \tag{3.108}$$

and

$$(-1)^{\nu_i} = \prod_{n_{j\neq i}=0,\pi; n_i=\pi} \delta(\Lambda_{n_1,n_2,n_3}), \tag{3.109}$$

for $i = 1, 2, 3$.

3.3 Topological insulators with inversion symmetry

So far, we have discussed systems with time-reversal symmetry. We will, however, also be interested in systems with other symmetries, and we will begin by studying systems with inversion symmetry. This is done in [5], and we will follow their approach. As described in section 2.4, the wave-function in the momentum representation transforms according to

$$\pi \left|\mathbf{k}, \sigma\right\rangle = \left|-\mathbf{k}, \sigma\right\rangle, \tag{3.110}$$

where σ denotes the spin of the system. This means that if we have a Hamiltonian of the form

$$H = \sum_{\mathbf{k},\sigma,\sigma'} \left|\mathbf{k}, \sigma\right\rangle H_{\sigma,\sigma'}(\mathbf{k}) \left\langle \mathbf{k}, \sigma'\right|, \tag{3.111}$$

it will transform according to

$$\pi H \pi^{-1} = \sum_{\mathbf{k},\sigma,\sigma'} \left|-\mathbf{k}, \sigma\right\rangle H_{\sigma,\sigma'}(\mathbf{k}) \left\langle -\mathbf{k}, \sigma'\right| = \sum_{\mathbf{k},\sigma,\sigma'} \left|\mathbf{k}, \sigma\right\rangle H_{\sigma,\sigma'}(-\mathbf{k}) \left\langle \mathbf{k}, \sigma'\right|. \tag{3.112}$$

Further assuming that

$$H(-\mathbf{k}) = H(\mathbf{k}), \tag{3.113}$$

we get that

$$\pi H \pi^{-1} = H, \tag{3.114}$$

which means that the system preserves inversion symmetry. The goal is now to express the $\mathbb{Z}_2$ invariant using the eigenvalues of π.

The Berry connection, as defined in equation (3.11), is a collection of three matrices, so the trace of the Berry connection is a vector. We denote this vector by $\mathbf{a}^C(\mathbf{k})$. Define the *Berry curvature* of $\mathbf{a}^C(\mathbf{k})$ by

$$F(\mathbf{k}) = \nabla_k \times \mathbf{a}^C(\mathbf{k}). \tag{3.115}$$

For a system that has time-reversal symmetry, one can show

$$F(-\mathbf{k}) = -F(\mathbf{k}). \tag{3.116}$$

If the system also has inversion symmetry, we also have

$$F(-\mathbf{k}) = F(\mathbf{k}). \tag{3.117}$$

This means that if the system has both time-reversal and inversion symmetry, then

$$F(\mathbf{k}) = 0 \tag{3.118}$$

for all $\mathbf{k}$, and thus that we can always choose a gauge such that

$$\mathbf{a}^C(\mathbf{k}) = 0. \tag{3.119}$$

To obtain such a gauge we begin by considering the matrix

$$v_{\alpha\beta}(\mathbf{k}) := \langle u_\alpha(\mathbf{k})|\pi\Theta|u_\beta(\mathbf{k})\rangle\,. \tag{3.120}$$

We can show that v is antisymmetric and unitary. Namely, we have

$$v_{\alpha\beta}(\mathbf{k}) = \langle \pi^\dagger u_\alpha(\mathbf{k})|\Theta|u_\beta(\mathbf{k})\rangle = \langle u_\alpha(-\mathbf{k})|\Theta|u_\beta(\mathbf{k})\rangle\,. \tag{3.121}$$

This is the same as the matrix w defined in equation (3.51). Since this matrix is unitary and antisymmetric at momenta where $\mathbf{k} = -\mathbf{k}$, we see that v must have these properties for all $\mathbf{k}$.

Furthermore, according to [21, eq. 3.4], we have

$$\frac{\mathrm{i}}{2}\mathrm{tr}\left[v^\dagger \nabla_k v\right] = \mathbf{a}^C(\mathbf{k}), \tag{3.122}$$

In [21] it is also argued that

$$\mathrm{tr}\left[v^\dagger \nabla_k v\right] = \nabla_k \mathrm{tr}\left[\log(v)\right] = \nabla_k \log\left[\det(v)\right]. \tag{3.123}$$

Inserting this into equation (3.122), we get

$$\mathbf{a}^C(\mathbf{k}) = \frac{\mathrm{i}\nabla_k}{2}\log(\det(v(\mathbf{k}))) = \mathrm{i}\nabla \log(\mathrm{Pf}\,[v(\mathbf{k})]). \tag{3.124}$$

This means that in order to obtain a gauge in which $\mathbf{a}^C(\mathbf{k}) = 0$, we have to adjust the phase of $|u_\alpha(\mathbf{k})\rangle$ so that $\mathrm{Pf}\,[v(\mathbf{k})] = 1$.

To obtain the $\mathbb{Z}_2$-invariant, we must calculate the w-matrix. We denote by $\xi_a(\Lambda_i) = \pm 1$, the eigenvalue of π for band a at Λ_i. This gives us

$$w_{\alpha\beta}(\Lambda_i) = \langle u_\alpha(-\Lambda_i)|\pi\pi\Theta|u_\beta(\Lambda_i)\rangle = \xi_\alpha(\Lambda_i)\langle u_\alpha(\Lambda_i)|\pi\Theta|u_\beta(\Lambda_i)\rangle = \xi_\alpha(\Lambda_i)v_{\alpha\beta}(\Lambda_i). \tag{3.125}$$

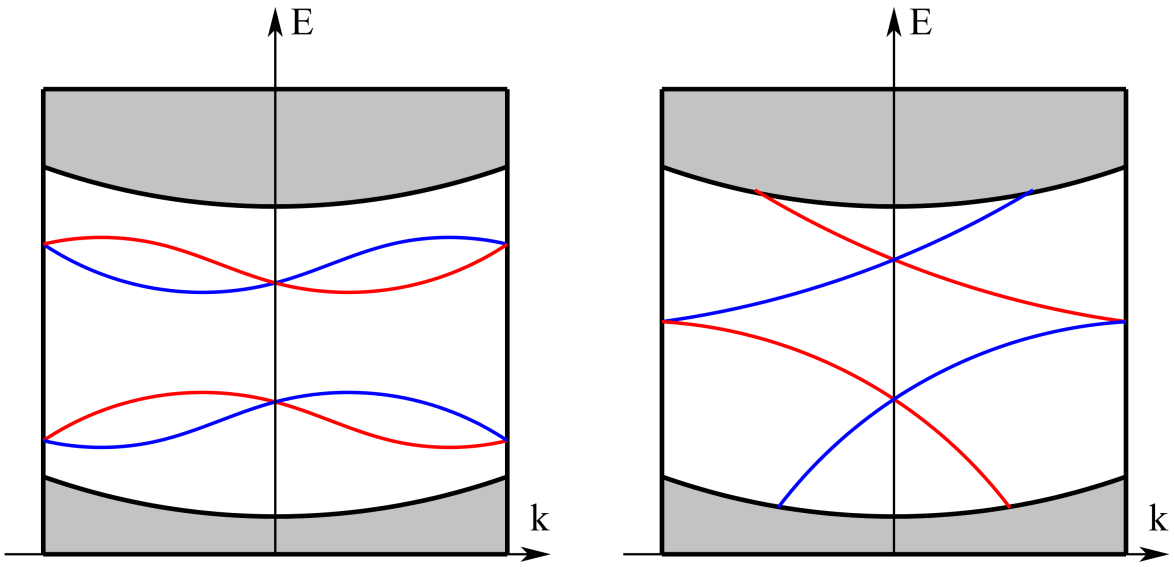

Figure 3: To the left we see an image of what the band structure in a trivial insulator could look like. We see surface states that reach into the bulk bandgap, but it is clear that the Fermi level could still be positioned so that it does not cross any energy level. To the right, we have an example of a topological insulator. Here we see that the Fermi level can never be positioned without crossing the surface states. We also note that the energy levels are degenerate at time-reversal invariant momenta. The grey areas mark the bulk energy bands.

We know that both the w-matrix and the v-matrix are antisymmetric. This means that $\xi_\alpha = \xi_\beta$ when $w_{\alpha\beta} \neq 0$. A non-zero $w_{\alpha\beta}$ is obtained only when the bands α and β form a Kramers pair. Therefore, if α and β are the nth Kramers pair in the total of $2N$ bands, we write $\xi_\alpha = \xi_\beta = \xi_{2n}$.

From equation (3.125) we see that

$$\mathrm{Pf}(w(\Lambda_i)) = \mathrm{Pf}(v(\Lambda_i)) \prod_{n=1}^{N} \xi_{2n}(\Lambda_i). \tag{3.126}$$

Choosing the gauge in which $\mathrm{Pf}\,[v(\mathbf{k})] = 1$, we get

$$\delta(\Lambda_i) = \frac{\mathrm{Pf}(w(\Lambda_i))}{\sqrt{\det(w(\Lambda_i))}} = \prod_{n=1}^{N} \xi_{2n}(\Lambda_i). \tag{3.127}$$

This means that the $\mathbb{Z}_2$ invariant can be calculated simply by using the eigenvalues of π at time-reversal invariant momenta Λ_i.

The physical consequence of a non-zero $\mathbb{Z}_2$ invariant is that we get topologically protected surface states. In this case, the surface states are guaranteed to cross the Fermi level, and we get a non-gapped system. An example of this is shown in Figure 3.

4 Topological crystalline insulators

The previous section concerned topological insulators. We showed how to find topological invariants for these. Now, these invariants depend on the existence of Kramers pairs, and as we saw, these pairs arise as a consequence of $\Theta^2 = -1$, i.e. in systems with half-integer spin. In systems with integer spin, which e.g occurs when the total number of fermions in a system is even, we have $\Theta^2 = 1$, which means that the reasoning in equations (3.46) and (3.47) does not guarantee that $|n\rangle$ and $\Theta |n\rangle$ are different states.

In order to find topological states in integer spin systems, we thus need something more than time-reversal symmetry alone. It turns out that a combination of time-reversal symmetry and crystal symmetry can give us what we need. Such a system, where we get topologically protected states because of this combination, is called a *topological crystalline insulator*. In fact, when evaluating the topological invariant used to classify ordinary topological insulators, the topological crystalline insulators fall into the category of trivial insulators [6, Ch. 3.1]. Thus to understand the behaviour of topological crystalline insulators, the previous considerations are insufficient.

In this section we will study topological crystalline insulators. In [8], these are described as materials which cannot be smoothly connected to a trivial atomic insulator when time-reversal symmetry and certain point group symmetries are respected. It is clear that a surface can break the symmetry of the bulk, depending on which surface we are considering. This means that there will only be some surfaces that support gapless surface states, i.e. those surfaces that respect the appropriate symmetry will support such states. Surfaces with low symmetry do not have robust surface states.

It is clear that there are a lot of different crystal symmetries that can be worth examining, and it lies beyond the scope of this thesis to describe this in full generality. Instead we will focus on the case where we have a system with integer spin and C_4-symmetry. This is done in [8], where it is shown that there exists a topological invariant that characterizes the band structure of a three-dimensional time-reversal invariant insulator with fourfold rotational symmetry. We denote this topological invariant by v_0 and will here follow [8] and argue that it is a $\mathbb{Z}_2$ invariant, i.e. can assume the values zero and one.

Consider the Bloch wave functions

$$|\psi_n(\mathbf{k})\rangle = e^{\mathrm{i}\mathbf{k}\bullet\mathbf{r}} |u_n(\mathbf{k})\rangle . \tag{4.1}$$

We let the unit cell be invariant under C_4 rotation around the z-axis. Together with time-reversal symmetry, this means that the Hamiltonian satisfies the following:

$$H(k_x, k_y, k_z) = RH(k_y, -k_x, k_z)R^{-1}, \tag{4.2}$$

and

$$H(\mathbf{k}) = \Theta H(-\mathbf{k})\Theta^{-1}, \tag{4.3}$$

where $R = e^{\mathrm{i}L_z\pi/2}$ is the C_4 rotation operator around the z-axis.

As already mentioned, for systems with integer spin we have

$$\Theta^2 = 1, \tag{4.4}$$

so time-reversal symmetry alone does not guarantee a twofold degeneracy of the energy bands. However, we will now show that together with the fourfold rotational symmetry, we can get protected degeneracies at four special momenta in the Brillouin zone,

$$\Gamma = (0,0,0),\ M = (\pi,\pi,0),\ Z = (0,0,\pi),\ A = (\pi,\pi,\pi). \tag{4.5}$$

At such a high symmetry point Λ_i, the Hamiltonian $H(\Lambda_i)$ commutes with R, which means that the energy eigenstates $|u_n(\Lambda_i)\rangle$ are eigenstates of fourfold rotation with eigenvalues ± 1 and $\pm i$. Also, since $H(\Lambda_i)$ is real, energy bands at Λ_i with $\pm i$ eigenvalues are guaranteed to be degenerate, forming a two-dimensional irreducible real representation of C_4. These degenerate energy bands can now be used in a way that is analogous to how the Kramers pairs are used. Namely, consider a set of energy bands that are doubly degenerate at Γ, M, A and Z. We denote the bands by

$$|u_n(\mathbf{k})\rangle\,, \quad n = 1, \ldots, 2N, \tag{4.6}$$

and let the corresponding energy eigenvalues be related by

$$E_{2n-1}(\Lambda_i) = E_{2n}(\Lambda_i). \tag{4.7}$$

The $\mathbb{Z}_2$ topological invariant is now defined similarly to the one for the toplogical insulators described in Section 3, namely

$$(-1)^{\nu_0} = (-1)^{\nu_{\Gamma M}}(-1)^{\nu_{AZ}}, \tag{4.8}$$

where $\nu_{\Gamma M}$ and ν_{AZ} are given by

$$(-1)^{\nu_{\Lambda_1\Lambda_2}} = \exp\left(\mathrm{i}\int_{\Lambda_1}^{\Lambda_2} d\mathbf{k} A(\mathbf{k})\right)\frac{\mathrm{Pf}\left[W(\Lambda_2)\right]}{\mathrm{Pf}\left[W(\Lambda_1)\right]}, \tag{4.9}$$

where in the case of $\nu_{\Gamma M}$, we integrate along an arbitrary line between Γ and M within the plane $k_z = 0$, and similarly, in the case of ν_{AZ}, we integrate along an arbitrary line between A and Z that lies in the plane $k_z = \pi$. The matrix $A(\mathbf{k})$ is as before the Berry connection matrix

$$A(\mathbf{k}) = -\mathrm{i}\sum_n \langle u_n(\mathbf{k})|\nabla_{\mathbf{k}}|u_n(\mathbf{k})\rangle\,, \tag{4.10}$$

and $W(\Lambda_i)$ is given by

$$W_{mn}(\Lambda_i) = \langle u_m(\mathbf{k})|R\Theta|u_n(\mathbf{k})\rangle\,. \tag{4.11}$$

We note that A and W both depend on the choice of basis for the system. It turns out though that $\nu_{\Lambda_1\Lambda_2}$ is gauge invariant. Namely, consider two different bases, where we denote the basis functions by $|u_n(\mathbf{k})\rangle$ and $|u'_n(\mathbf{k})\rangle$ respectively. These bases are related to each other by a gauge transformation

$$|u'_n(\mathbf{k})\rangle = \sum_m G_{nm}(\mathbf{k})\,|u_m(\mathbf{k})\rangle\,, \tag{4.12}$$

where $G \in U(2N)$. Using this gauge transformation together with the fact that $\mathrm{Pf}(X^T M X) = \det(X)\mathrm{Pf}(M)$, one can show that indeed

$$\tilde{\nu}_{\Lambda_1\Lambda_2} = \nu_{\Lambda_1\Lambda_2}. \tag{4.13}$$

Now we want to argue that $\nu_{\Lambda_1\Lambda_2}$ is a $\mathbb{Z}_2$ invariant, i.e. that it assumes the values 0 and 1, possibly mod 2. We begin by noting that at the planes $k_z = 0$ and $k_z = \pi$, we have

$$H(\mathbf{k}) = R^2\Theta H(\mathbf{k})(R^2\Theta)^{-1}. \tag{4.14}$$

This follows from Equations (4.2) and (4.3), together with the periodicity of the material, which implies that $H(\pi) = H(-\pi)$. Because of this, we can use the following real gauge for evaluation of the integral in equation (4.9):

$$R^2\Theta \left|u_n(\mathbf{k})\right\rangle = -\left|u_n(\mathbf{k})\right\rangle. \tag{4.15}$$

We know that Θ is an anti-unitary operator, therefore also $R^2\Theta$ must be anti-unitary. In particular, this means that A is zero everywhere along the integration paths that we use in the two integrals. This means that

$$(-1)^{\nu_{\Lambda_1,\Lambda_2}} = \frac{\mathrm{Pf}(w(\Lambda_2))}{\mathrm{Pf}(w(\Lambda_1))}. \tag{4.16}$$

Now, one can argue that

$$R^2 \left|u_n(\Lambda_i)\right\rangle = -\left|u_n(\Lambda_i)\right\rangle, \tag{4.17}$$

which together with equation (4.14) gives us that

$$\Theta \left|u_n(\Lambda_i)\right\rangle = \left|u_n(\Lambda_i)\right\rangle. \tag{4.18}$$

This means that the wave-function is real, and also that we have

$$w_{mn}(\Lambda_i) = \langle u_m(\mathbf{k})|R|u_n(\mathbf{k})\rangle. \tag{4.19}$$

One can argue that it is now possible to choose the real basis

$$\left|u_{2m}(\Lambda_i)\right\rangle = R\left|u_{2m-1}(\Lambda_i)\right\rangle. \tag{4.20}$$

This means that $w(\Lambda_i)$ will become a direct sum of N two-by-two Levi-Civita tensors. We denote this special case by w^0. In a more general case, w can be written as

$$w(\Lambda_i) = O^T(\Lambda_i)w^0O(\Lambda_i), \tag{4.21}$$

where $O(\Lambda_i)$ is an orthogonal $2N \times 2N$ matrix.

This means that

$$\mathrm{Pf}(w(\Lambda_i)) = \mathrm{Pf}(O^T(\Lambda_i)w^0O(\Lambda_i)) = \det(O(\Lambda_i))\mathrm{Pf}(w^0). \tag{4.22}$$

The Pfaffian of w^0 is 1, which leaves us with

$$\mathrm{Pf}(w(\Lambda_i)) = \pm 1. \tag{4.23}$$

So

$$(-1)^{\nu_{\Lambda_1\Lambda_2}} = \frac{\mathrm{Pf}(w(\Lambda_2))}{\mathrm{Pf}(w(\Lambda_1))} = \pm 1, \tag{4.24}$$

and therefore $\nu_{\Lambda_1\Lambda_2}$ must be either 0 or 1.

We now note from equation (4.8) that also ν_0 must be 1 or 0 modulo 2. The interpretation of these three topological invariants is the following. The invariants $\nu_{\Lambda_1\Lambda_2}$ characterize the band structures in two dimensions in the planes $k_z = 0$ and $k_z = \pi$, while the invariant ν_0 characterizes the band-structure in three dimensions and tells us that there exist gapless surface states on the (001)-surface only when $\nu_0 = 1$.

5 The Slater-Koster tight-binding rules

In previous sections we have seen a theoretical description of topological phases in materials. We have seen that the topological phases can be recognized by gapless surface states in otherwise insulating materials. Thus, it is of interest to study the band structures of such materials. This is commonly done using tight-binding calculations, which is a method that is based on the linear combination of atomic orbitals (LCAO) method used for molecules.

Before we get into the details of how this approximation is done, we will first describe the problem that we actually want to solve. To do this, we follow what is done in [22] and [23].

We want to solve the Schrödinger equation for the electrons in a crystal, i.e.

$$H\left|\psi\right\rangle = E\left|\psi\right\rangle. \tag{5.1}$$

According to Bloch's theorem the wave-functions obey

$$\left|\psi(\mathbf{k},\mathbf{r})\right\rangle = \exp(\mathrm{i}\mathbf{k}\bullet\mathbf{r})\left|u_{\mathbf{k}}(\mathbf{r})\right\rangle, \tag{5.2}$$

where $u_{\mathbf{k}}(\mathbf{r})$ is a function that has the periodicity of the lattice. For realistic potentials it is impossible to find exact solutions of equation (5.1), so an approximation is needed.

In the *tight-binding model* one assumes that it is possible to construct Bloch waves from atomic orbitals, belonging to the individual atoms in the lattice. We denote these atomic orbitals by $\phi_n(\mathbf{r}-\mathbf{R})$, where $\mathbf{R}$ tells us the position of the atom in the lattice n indicates that it belongs to the energy eigenvalue E_n.

Now Bloch basis functions are defined by

$$\left|\phi_n(\mathbf{k},\mathbf{r})\right\rangle := \frac{1}{\sqrt{N}}\sum_{\mathbf{R}}\exp(\mathrm{i}\mathbf{k}\bullet\mathbf{R})\left|\phi_n(\mathbf{r}-\mathbf{R})\right\rangle, \tag{5.3}$$

where $\mathbf{R}$ denotes positions in the lattice.

These basis functions are used to approximate the solutions of (5.1) by

$$\left|\psi(\mathbf{k},\mathbf{r})\right\rangle = \sum_n c_n(\mathbf{k})\left|\phi_n(\mathbf{k},\mathbf{r})\right\rangle, \tag{5.4}$$

where $c_n(\mathbf{k})$ are the constants we want to determine.

The atomic orbitals are not necessarily orthogonal to each other, which means that we need to use the *overlap matrix* of the atomic orbitals defined by

$$S_{jl} = \langle\phi_m(\mathbf{r}-\mathbf{R}_j)|\phi_n(\mathbf{r}-\mathbf{R}_l)\rangle. \tag{5.5}$$

Finding the coefficients $c_i(\mathbf{k})$, defined in equation (5.4), corresponds to solving the generalized eigenvalue problem

$$HC = ESC, \tag{5.6}$$

where H is the Hamiltonian matrix

$$H_{mn} = \langle \phi_m(\mathbf{k}, \mathbf{r}) | H | \phi_n(\mathbf{k}, \mathbf{r}) \rangle \,, \tag{5.7}$$

and C is a vector containing the coefficients c_i. Expressing H in terms of the atomic orbitals, one gets

$$H_{mn} = \sum_{\mathbf{R}} \exp(\mathrm{i}\mathbf{k} \bullet \mathbf{R}) \int d^3r \, \phi_m(\mathbf{r} - \mathbf{R}) H \phi_n(\mathbf{r}) \tag{5.8}$$

There are several difficulties that arise when trying to solve equation (5.6) numerically. First of all, one needs to calculate the Hamiltonain matrix and the overlap matrix. This in itself is difficult since it involves integrals of atomic orbitals. Then there is also the problem of the huge number of atoms in a system. It is not possible to take all such interactions into account. In what follows we will see how one can compute a new orthogonal basis from the atomic orbitals, and how one can approximate the Hamiltonian matrix elements in a suitable way.

5.1 The Slater-Koster rules

In this section, we will study the tight-binding theory developed by Slater and Koster in [9]. As mentioned, one problem of solving equation (5.6) is that the overlap matrix is not the identity matrix, so we do not get an ordinary eigenvalue problem. To remedy this, we use the method described by Löwdin in [24] to create an orthogonal basis of atomic wave functions, and get the following new atomic orbitals, which we call *Löwdin functions*:

$$\psi_n = \sum_m \phi_m S_{mn}^{-\frac{1}{2}}, \tag{5.9}$$

where S is the overlap matrix defined in equation (5.5). The square root of S is in this case chosen to be the principal square root of S. This is calculated in the following way: First, diagonalize S, using a unitary matrix U, so we get

$$S = UDU^\dagger. \tag{5.10}$$

This can be done since the Hamiltonian matrix is Hermitian. Now, the overlap matrix is a positive definite matrix [25], which means that D is a diagonal matrix with positive elements on the diagonal. Defining $D^{1/2}$ as the diagonal matrix whose diagonal elements are the positive square roots of the elements in D, we define the *principal square root* of S by

$$S^{1/2} = UD^{1/2}U^\dagger. \tag{5.11}$$

Similarly, we have

$$S^{-1/2} = UD^{-1/2}U^\dagger, \tag{5.12}$$

where $D^{-1/2}$ is the inverse of $D^{1/2}$.

That the Löwdin functions are orthogonal to each other is easily seen, [26, sec. 9.1.3.1], by noting that

$$\begin{aligned} \langle \psi_n | \psi_m \rangle &= \langle \sum_k \phi_k S_{kn}^{-1/2} | \sum_l \phi_l S_{lm}^{-1/2} \rangle = \sum_{kl} \left(S_{kn}^{-1/2} \right)^* S_{lm}^{-1/2} \langle \phi_k | \phi_l \rangle \\ &= \sum_{kl} S_{kn}^{-1/2} S_{lm}^{-1/2} S_{kl}, \end{aligned} \tag{5.13}$$

where in the last equality we have used that S is real, which is true in case the wave functions are real [27]. This now gives us

$$\langle \psi_n | \psi_m \rangle = \delta_{nm}, \tag{5.14}$$

and thus the Löwdin functions are orthogonal.

From [9] we now have the following important result, which tells us that the same symmetry arguments can be used both for the original atomic orbitals and the Löwdin functions.

Proposition 5.1. *Consider the sets of atomic orbitals $\{\phi_n\}$ and Löwdin functions $\{\psi_n\}$, and assume that the functions are real and that invariance of S under a unitary transformation implies invariance of $S^{-1/2}$ under the same transformation. Let G be the symmetry group of a crystal and let $g \in G$. Then there is a labelling of the functions in the respective sets such that*

$$g.\phi_n = \sum_m \phi_m \Gamma(g)_{mn} \Leftrightarrow g.\psi_n = \sum_m \psi_m \Gamma(g)_{mn}, \tag{5.15}$$

where $\Gamma(g)$ is a unitary matrix.

Proof. To prove the statement, we need to show that the transformation properties of the Löwdin functions are the same as those of the ordinary atomic orbitals. Let G be the symmetry group of the crystal we are considering, and let g be an element in G. There is a unitary representation of this group. Then, suppose that, for an atomic orbital ϕ_n, we have

$$g.\phi_n = \sum_m \phi_m \Gamma(g)_{mn}, \tag{5.16}$$

where $\Gamma(g)$ is a unitary matrix representing g. We know from earlier that the Löwdin functions are related to the atomic orbitals by

$$\psi_n = \sum_m \phi_m S^{-1/2}_{mn}. \tag{5.17}$$

Letting g act on ψ_n, gives us

$$g.\psi_n = g.\sum_m \phi_m S^{-1/2}_{mn} = \sum_m g.\phi_m S^{-1/2}_{mn} = \sum_m \sum_l \phi_l \Gamma(g)_{lm} S^{-1/2}_{mn}. \tag{5.18}$$

Now, consider the overlap matrix. We have

$$S_{mn} = \int \phi^*_m \phi_n d\tau = \int (g.\phi_m)^* (g.\phi_n) d\tau. \tag{5.19}$$

Inserting the expression for the transformation given in equation (5.16), we have

$$\begin{aligned} S_{mn} &= \sum_k \sum_l \Gamma(g)^*_{km} \Gamma(g)_{ln} \int \phi^*_k \phi_l d\tau = \sum_k \sum_l \Gamma(g)^*_{km} \Gamma(g)_{ln} S_{kl} \\ &= \sum_k \sum_l \Gamma(g)^*_{km} S_{kl} \Gamma(g)_{ln}. \end{aligned} \tag{5.20}$$

So the overlap matrix is invariant with respect to the unitary transformation. This in turn means that the inverse of the square root of S is invariant under this transformation, which can be written as

$$S^{-1/2} = \Gamma(g) S^{-1/2} \Gamma(g)^{-1}. \tag{5.21}$$

Now, we get

$$g.\psi_n = \sum_m \sum_l \phi_l \Gamma(g)_{lm} S^{-1/2}_{mn}, \tag{5.22}$$

which in turn gives us

$$g.\psi_n = \sum_m \sum_l \phi_l S^{-1/2}_{lm} \Gamma(g)_{mn}, \tag{5.23}$$

since $S^{-1/2}$ is invariant under g. But now we note that by definition of ψ, we have

$$g.\psi_n = \sum_m \psi_m \Gamma(g)_{mn}, \tag{5.24}$$

and thus the Löwdin functions transform in the same way as the atomic orbitals under the symmetry group of the crystal. □

What the proof says is that given an atomic orbital ϕ_n that transforms according to

$$g.\phi_n = \sum_m \phi_m \Gamma(g)_{mn}, \tag{5.25}$$

there is a corresponding Löwdin function, ψ_n, that transforms in the same way, i.e.

$$\psi_n = \sum_m \psi_m \Gamma(g)_{mn}. \tag{5.26}$$

This means that the Löwdin functions have the same symmetry properties as the original atomic orbitals, so we can use the same symmetry arguments in both cases. Also, this means that we can use the same notation for the Löwdin functions as we do for the atomic orbitals, i.e. s, p, d and their variants without mathematical complications.

Assuming periodic boundary conditions with N being the number of unit cells in the repeating region, we create the following Bloch sums of the Löwdin orbitals,

$$\Psi_n(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_i \exp(\mathrm{i}\mathbf{k} \bullet \mathbf{R}_i) \psi_n(\mathbf{r} - \mathbf{R}_i), \tag{5.27}$$

where $\mathbf{R}_i$ is the position of the atom on which the orbital is located. These sums are orthonormal.

We want to express the Hamiltonian matrix using the Bloch sums, which we do in the following way:

$$H_{mn}(\mathbf{r}) = \int \Psi^*_m(\mathbf{r}) H \Psi_n(\mathbf{r}) = \frac{1}{N} \sum_{i,j} \exp\left[\mathrm{i}\mathbf{k} \bullet (\mathbf{R}_j - \mathbf{R}_i)\right] \int \psi^*_m(\mathbf{r} - \mathbf{R}_i) H \psi_n(\mathbf{r} - \mathbf{R}_j)\, dv. \tag{5.28}$$

The orbitals ψ_m are located at positions $\mathbf{R}_i$ in the unit cell, while the orbitals ψ_n are located at positions $\mathbf{R}_j$. The positions $\mathbf{R}_i$ and $\mathbf{R}_j$ are not necessarily the same. Despite this, we note that one of the sums can be eliminated, since the only effect it has is multiplying the other sum by N. This gives us

$$H_{mn} = \sum_j \exp\left[\mathrm{i}\mathbf{k} \bullet (\mathbf{R}_j - \mathbf{R}_{i_0})\right] \int \psi^*_m(\mathbf{r} - \mathbf{R}_{i_0}) H \psi_n(\mathbf{r} - \mathbf{R}_j)\, dv, \tag{5.29}$$

for some i_0. In addition to this, we can choose our coordinates in such a way that $\mathbf{R}_{i_0} = 0$, thus

$$H_{mn} = \sum_j \exp\left[\mathrm{i}\mathbf{k} \bullet (\mathbf{R}_j)\right] \int \psi_m^*(\mathbf{r}) H \psi_n(\mathbf{r} - \mathbf{R}_j)\, dv, \tag{5.30}$$

which gives us back the expression we had in equation (5.8), but with a different set of orbitals. We note that each term in this expression is associated with a pair of orbitals on neighbouring (not only nearest neighbours) atoms.

In order to evaluate the Hamiltonian matrix elements, we need to calculate the integrals that arise. This is in general very hard, and one needs to make some sort of approximation. In [9], the method is to approximate the integrals by fitting them to accurate determinations of the energy at certain suitable **k**-points.

Clearly, from equation (5.30), there is a large number of integrals that needs to be calculated. The first thing to note is that for symmetry reasons many of the integrals are equal, so one does not need to calculate all of them. In order to simplify the problem further, it is reasonable to see if it is possible to neglect some of the terms. Indeed, as is discussed in [9], this is the case.

The first thing to note, is that the integrals in (5.30), will get smaller the larger $|\mathbf{R}_j|$ is. The decrease will not be as rapid as for the original atomic orbitals, but it is still reasonable to assume that we can neglect all integrals for which the interatomic distance is larger than some set value. It is common, and often sufficient, to take only nearest and next-nearest neighbour interactions into account, but depending on the desired accuracy it is possible to use more.

Another approximation we make is to neglect three-center integrals. This approximation is usually called the *two-center approximation*, and is e.g. described in [28]. Each Hamiltonian matrix element consists of an integral of three functions, one potential function and two orbital functions, centered at three atomic sites. If all these three functions are centered at the same site, this is called an *on-site* matrix element. If the orbitals are at different sites, but within the distance we are considering, and the potential is at one of those two sites, the integral is called a two-center integral, or a *hopping integral*. The other cases, where the distance between the atomic orbitals is large, or where the three functions are centered at three different sites (i.e. *three-center integrals*) are ignored.

Neglecting the three-center integrals, we see that the integrals we have left become similar to those for diatomic molecules. The vector $\mathbf{R}_j$ stretching from one atom to another, can be considered as being the axis in a diatomic molecule. We can thus express each of the Löwdin functions ψ as a sum of functions that are space quantized with respect to that axis. Since the Löwdin functions have the same symmetry properties as the atomic orbitals, we can make the approximation that we can expand the Löwdin functions in the same way as the atomic orbitals. This means that if, for example, ψ corresponds to an atomic p orbital, we can express it as a linear combination of $p\sigma$ and $p\pi_\pm$ functions with respect to the axis $\mathbf{R}_j$. In the integral (5.30), we now get a non-zero contribution only if we are dealing with the same type of component (i.e. σ or $\pi_\pm$ etc.) of both ψ_n and ψ_m. Thus we get a relatively small number of integrals in (5.30).

Now we wish to describe the integrals that remain. To do this, we set up the atomic orbitals with respect to a set of rectangular axes. We symbolize the p_x, p_y and p_z functions by x, y and z respectively and similarly we symbolize the various d functions by $x^2 - y^2, 3z^2 - r^2, xy, yz$ and zx. Let the direction cosines for the vector $\mathbf{R}_j$, i.e. the axis that we are considering, be (l, m, n).

Now we denote the integral in which the ψ_α function is of type α and ψ_β is of type β with direction cosines (l, m, n) by

$$E_{\alpha,\beta}(l, m, n) = \int \psi^*_\alpha(\mathbf{r}) H \psi_\beta(\mathbf{r} - \mathbf{R}_j) dv. \tag{5.31}$$

Since we can expand the Löwdin functions in terms of the space quantized functions mentioned above, these integrals can be split into linear combinations of other integrals. For example, $E_{x,xy}(l, m, n)$ can (approximately) be written as a sum of two integrals; that between a $p\sigma$ orbital on the first atom and a $d\sigma$ orbital on the second atom and that between a $p\pi$ orbital on the first atom and a $d\pi$ orbital on the second one. We denote these integrals by $(pd\sigma)$ and $(pd\pi)$ respectively. A closer examination reveals that

$$E_{x,xy}(l, m, n) = \sqrt{3} l^2 m (pd\sigma) + m(1 - 2l^2)(pd\pi). \tag{5.32}$$

The rest of all such integrals are listed in Table 1.

The integrals $(\alpha\beta\gamma)$ are all functions of distance, but since we are considering interactions only at fixed distances, we will denote the different distances by a lower index, as can be seen in Table 2.

Now, the integrals $E_{\alpha,\beta}(k, l, m)$ are not the Hamiltonian matrix elements, these are given by equation (5.30). In order to calculate these matrix elements, we therefore have to combine several of these E-integrals. This can easily be calculated, and the result for several different common crystal structures is described in [9]. For example, the results are tabulated for cubic crystals with one kind of atom. Since we are interested in rock-salt crystal structures, the simple cubic structure is of relevance here, and we describe some of the results.

Table 1: Table over the integrals $E_{\alpha\beta}(k,l,m)$

$E_{s,s}$	$(ss\sigma)$
$E_{s,x}$	$l(sp\sigma)$
$E_{x,x}$	$l^2(pp\sigma)+(1-l^2)(pp\pi)$
$E_{x,y}$	$lm(pp\sigma)-lm(pp\pi)$
$E_{x,z}$	$ln(pp\sigma)-ln(pp\pi)$
$E_{s,xy}$	$\sqrt{3}lm(sd\sigma)$
E_{s,x^2-y^2}	$\frac{\sqrt{3}}{2}(l^2-m^2)(sd\sigma)$
$E_{s,3z^2-r^2}$	$\left[n^2-\frac{1}{2}(l^2+m^2)\right](sd\sigma)$
$E_{x,xy}$	$\sqrt{3}l^2m(pd\sigma)+m(1-2l^2)(pd\pi)$
$E_{x,yz}$	$\sqrt{3}lmn(pd\sigma)-2lmn(pd\pi)$
$E_{x,zx}$	$\sqrt{3}l^2n(pd\sigma)+n(1-2l^2)(pd\pi)$
E_{x,x^2-y^2}	$\frac{\sqrt{3}}{2}l(l^2-m^2)(pd\sigma)+l(1-l^2+m^2)(pd\pi)$
E_{y,x^2-y^2}	$\frac{\sqrt{3}}{2}m(l^2-m^2)(pd\sigma)-m(1+l^2-m^2)(pd\pi)$
E_{z,x^2-y^2}	$\frac{\sqrt{3}}{2}n(l^2-m^2)(pd\sigma)-n(l^2-m^2)(pd\pi)$
$E_{x,3z^2-r^2}$	$l\left[n^2-\frac{1}{2}(l^2-m^2)\right](pd\sigma)-\sqrt{3}ln^2(pd\pi)$
$E_{y,3z^2-r^2}$	$m\left[n^2-\frac{1}{2}(l^2-m^2)\right](pd\sigma)-\sqrt{3}mn^2(pd\pi)$
$E_{z,3z^2-r^2}$	$n\left[n^2-\frac{1}{2}(l^2-m^2)\right](pd\sigma)-\sqrt{3}n(l^2+m^2)(pd\pi)$
$E_{xy,xy}$	$3l^2m^2(dd\sigma)+(l^2+m^2-4l^2m^2)(dd\pi)+(n^2+l^2m^2)(dd\delta)$
$E_{xy,yz}$	$3lm^2n(dd\sigma)+ln(1-4m^2)(dd\pi)+ln(m^2-1)(dd\delta)$
$E_{xy,zx}$	$3l^2mn(dd\sigma)+mn(1-4l^2)(dd\pi)+mn(l^2-1)(dd\delta)$
E_{xy,x^2-y^2}	$\frac{3}{2}lm(l^2-m^2)(dd\sigma)+2lm(m^2-l^2)(dd\pi)+\frac{1}{2}lm(l^2-m^2)(dd\delta)$
E_{yz,x^2-y^2}	$\frac{3}{2}mn(l^2-m^2)(dd\sigma)-mn\left[1+2(l^2-m^2)\right](dd\pi)+mn\left[1+\frac{1}{2}(l^2-m^2)\right](dd\delta)$
E_{zx,x^2-y^2}	$\frac{3}{2}nl(l^2-m^2)(dd\sigma)+nl\left[1-2(l^2-m^2)\right](dd\pi)-nl\left[1-\frac{1}{2}(l^2-m^2)\right](dd\delta)$
$E_{xy,3z^2-r^2}$	$\sqrt{3}lm\left[n^2-\frac{1}{2}(l^2-m^2)\right](dd\sigma)-2\sqrt{3}lmn^2(dd\pi)+\frac{\sqrt{3}}{2}lm(1+n^2)(dd\delta)$
$E_{yz,3z^2-r^2}$	$\sqrt{3}mn\left[n^2-\frac{1}{2}(l^2+m^2)\right](dd\sigma)-\sqrt{3}mn(l^2+m^2-n^2)(dd\pi)-\frac{\sqrt{3}}{2}mn(l^2+m^2)(dd\delta)$
$E_{zx,3z^2-r^2}$	$\sqrt{3}ln\left[n^2-\frac{1}{2}(l^2+m^2)\right](dd\sigma)-\sqrt{3}ln(l^2+m^2-n^2)(dd\pi)-\frac{\sqrt{3}}{2}ln(l^2+m^2)(dd\delta)$
$E_{x^2-y^2,x^2-y^2}$	$\frac{3}{4}(l^2-m^2)^2(dd\sigma)+\left[l^2+m^2-(l^2-m^2)^2\right](dd\pi)+\left[n^2+\frac{1}{4}(l^2-m^2)^2\right](dd\delta)$
$E_{x^2-y^2,3z^2-r^2}$	$\frac{\sqrt{3}}{2}(l^2-m^2)\left[n^2-\frac{1}{2}(l^2+m^2)\right](dd\sigma)+\sqrt{3}n^2(m^2-l^2)(dd\pi)+\frac{\sqrt{3}}{4}(1+n^2)(l^2-m^2)(dd\delta)$
$E_{3z^2-r^2,3z^2-r^2}$	$\left[n^2-\frac{1}{2}(l^2+m^2)\right]^2(dd\sigma)+3n^2(l^2+m^2)(dd\pi)+\frac{3}{4}(l^2+m^2)^2(dd\delta)$

5.2 The simple cubic structure

Now we explicitly show how the Slater-Koster rules can be used in a special case, namely the simple cubic structure. To begin with, we take only nearest neighbour interactions into account. Let a be the lattice constant, and let the atoms in the crystal be located at points

$$\mathbf{R} = (pa, qa, ra). \tag{5.33}$$

If we rewrite the E-integrals in terms of p, q, r instead of the direction cosines, we get that

$$E_{\alpha,\beta}(p,q,r) = \int \psi_\alpha^*(\mathbf{r}) H \psi_\beta(\mathbf{r} - \mathbf{R}). \tag{5.34}$$

We have that p, q, r are related to l, m, n by

$$l = \frac{p}{\sqrt{p^2+q^2+r^2}}, \tag{5.35}$$

$$m = \frac{q}{\sqrt{p^2+q^2+r^2}}, \tag{5.36}$$

$$n = \frac{r}{\sqrt{p^2+q^2+r^2}}. \tag{5.37}$$

The nearest neighbours in the simple cubic structure have positions $\mathbf{R} = (\pm 1, 0, 0)$, $(0, \pm 1, 0)$ and $(0, 0, \pm 1)$. We here give an example of how to calculate a Hamiltonian matrix element:

$$\begin{aligned}(x/xy) = \sum_{\mathbf{R}} E_{x,xy}(l,m,n) &= E(0,0,0) + \exp(\mathrm{i}ak_x) E_{x,xy}(1,0,0) \\ &+ \exp(-\mathrm{i}ak_x) E_{x,xy}(-1,0,0) + \exp(\mathrm{i}ak_y) E_{x,xy}(0,1,0) \\ &+ \exp(-\mathrm{i}ak_y) E_{x,xy}(0,-1,0) + \exp(\mathrm{i}ak_z) E_{x,xy}(0,0,1) \\ &+ \exp(-\mathrm{i}ak_z) E_{x,xy}(0,0,-1).\end{aligned} \tag{5.38}$$

From Table 1 we have

$$E_{x,xy}(l,m,n) = \sqrt{3} l^2 m (pd\sigma) + m(1-2l^2)(pd\pi), \tag{5.39}$$

which means

$$E_{x,xy}(1,0,0) = E_{x,xy}(-1,0,0) = E_{x,xy}(0,0,1) = E_{x,xy}(0,0,-1) = 0, \tag{5.40}$$

and

$$E_{x,xy}(0,1,0) = -E_{x,xy}(0,-1,0) = (pd\pi). \tag{5.41}$$

Inserting this into equation (5.38) we get

$$(x/xy) = \exp(\mathrm{i}ak_y)(pd\pi) - \exp(-\mathrm{i}ak_y)(pd\pi) = 2\mathrm{i}\sin(ak_y)(pd\pi). \tag{5.42}$$

The rest of the matrix elements can be calculated in a similar fashion, and the result, which is described for up to second nearest neighbour interactions in [9], is listed in Table 2.

Table 2: Hamiltoninan matrix elements. For notational simplicity, we let $\xi = ax$, $\eta = ay$ and $\zeta = az$. The matrix elements are given up to next-nearest neighbours.

(s/s)	$s_0 + 2(ss\sigma)_1(\cos\xi + \cos\eta + \cos\zeta) + 4(ss\sigma)_2(\cos\xi\cos\eta + \cos\xi\cos\zeta + \cos\eta\cos\zeta)$
(s/x)	$2\mathrm{i}(sp\sigma)_1 \sin\xi + 2\sqrt{2}\mathrm{i}(sp\sigma)_2(\sin\xi\cos\eta + \sin\xi\cos\zeta)$
(s/xy)	$-2\sqrt{3}(sp\sigma)_2 \sin\xi\sin\eta$
(s/x^2-y^2)	$\sqrt{3}(sd\sigma)_1(\cos\xi - \cos\eta) + \sqrt{3}(sd\sigma)_2(\cos\xi\cos\zeta + \cos\eta\cos\zeta)$
$(s/3z^2-r^2)$	$(sd\sigma)_1(-\cos\xi - \cos\eta + 2\cos\zeta) + (sd\sigma)_2(-2\cos\xi\cos\eta + \cos\xi\cos\zeta + \cos\eta\cos\zeta)$
(x/x)	$p_0 + 2(pp\sigma)_1\cos\xi + 2(pp\pi)_1(\cos\eta + \cos\zeta) + 2(pp\sigma)_2(\cos\xi\cos\eta + \cos\xi\cos\zeta)$ $+2(pp\pi)_2(\cos\xi\cos\eta + \cos\xi\cos\zeta + 2\cos\eta\cos\zeta)$
(x/y)	$-2\left[(pp\sigma)_2 - (pp\pi)_2\right]\sin\xi\sin\eta$
(x/xy)	$2\mathrm{i}(pd\pi)_1 + \sqrt{6}\mathrm{i}(pd\sigma)_2\cos\xi\sin\eta + 2\sqrt{2}\mathrm{i}(pd\pi)_2\sin\eta\cos\zeta$
(x/yz)	0
(x/x^2-y^2)	$\sqrt{3}(pd\sigma)_1\sin\xi - \sqrt{\frac{3}{2}}(pd\sigma_2)\mathrm{i}\sin\xi\cos\zeta + \sqrt{2}(pd\pi)_2\mathrm{i}(2\sin\xi\cos)$
$(x/3z^2-r^2)$	$-(pd\sigma)_1\mathrm{i}\sin\xi + \sqrt{2}(pd\sigma)_2\mathrm{i}\left[\sin\xi\cos\zeta - \frac{1}{2}\sin\xi\cos\zeta\right] - \sqrt{6}(pd\pi)_2\mathrm{i}\sin\xi\cos\zeta$
$(z/3z^2-r^2)$	$2\mathrm{i}(pd\sigma)_1\sin\zeta + \mathrm{i}\left[\frac{1}{\sqrt{2}}(pd\sigma)_2 + \sqrt{6}(pd\pi)_2\right]\left[\cos\xi\sin\zeta + \cos\eta\sin\zeta\right]$
(xy/xy)	$d_0 + 2(dd\pi)_2(\cos\xi + \cos\eta) + 2(dd)_1\cos\zeta + 3(dd\sigma)_2\cos\xi\cos\eta$ $+2(dd\pi)_2(\cos\xi\cos\zeta + \cos\eta\zeta) + (dd\delta)_2(\cos\xi\cos\eta + 2\cos\xi\cos\zeta + 2\cos\eta\cos\zeta)$
(xy/xz)	$2\left[-(dd\pi)_2 + (dd\delta)_2\right]\sin\eta\sin\zeta$
(xy/x^2-y^2)	0
$(xy/3z^2-r^2)$	$\sqrt{3}\left[(dd\sigma)_2 - (dd\delta)_2\right]\sin\xi\sin\eta$
(xz/x^2-y^2)	$-\frac{3}{2}\left[(dd\sigma)_2 - (dd\delta)_2\right]\sin\xi\sin\zeta$
$(xz/3z^2-r^2)$	$-\frac{\sqrt{3}}{2}\left[(dd\sigma)_2 - (dd\delta)_2\right]\sin\xi\sin\zeta$
(x^2-y^2/x^2-y^2)	$d_0 + \frac{3}{2}(dd\sigma)_1(\cos\xi + \cos\eta) + (dd\delta)_1(\frac{1}{2}\cos\xi + \frac{1}{2}\cos\eta + 2\cos\zeta)$ $+4(dd\pi)_2\cos\xi\cos\eta + \left[\frac{3}{4}(dd\sigma)_2 + (dd\pi)_2 + \frac{9}{4}(dd\delta)_2\right](\cos\xi\cos\zeta + \cos\eta\cos\zeta)$
$(3z^2-r^2/3z^2-r^2)$	$d_0 + (dd\sigma)_1(\frac{1}{2}\cos\xi + \frac{1}{2}\cos\eta + 2\cos\zeta) + \frac{3}{2}(dd\delta)_1(\cos\xi + \cos\eta) +$ $(dd\sigma)_2(\cos\xi\cos\eta + \frac{1}{4}\cos\xi\cos\zeta + \frac{1}{4}\cos\eta\cos\zeta) + 3(dd\pi)_2(\cos\xi\cos\zeta + \cos\eta\cos\zeta)$ $+3(dd\delta)_2(\cos\xi\cos\eta + \frac{1}{4}\cos\xi\cos\zeta + \frac{1}{4}\cos\eta\cos\zeta)$
$(x^2-y^2/3z^2-r^2)$	$\frac{\sqrt{3}}{2}\left[-(dd\sigma)_1 + (dd\delta)_1\right](\cos\xi - \cos\eta) +$ $\left[\frac{\sqrt{3}}{4}(dd\sigma)_2 - \sqrt{3}(dd\pi)_2 + \frac{3\sqrt{3}}{4}(dd\delta)_2\right](\cos\xi\cos\zeta - \cos\eta\cos\zeta)$

We have not listed all matrix elements in Table 2, because all other elements can be derived using the following two symmetry-arguments:

1. Interchanging the order of the two orbitals in an integral changes the sign of the integral if the sum of the parities of the orbitals is odd. If the sum of the parities is even, nothing happens.
2. Cyclically permuting the coordinates and direction cosines does not change the value of the integral.

In order to be able to use this, we need to know which parities the different orbitals have. This is very simple, s- and d-orbitals have parity 0, while p-orbitals have parity 1. This means that the only situation in which we get a change of sign is when we swap a p-orbital with either an s- or a d-orbital.

For example we have

$$(x/xy) = (y/yz) = (z/zx) = -(xy/x) = -(yz/y) = -(zx/z). \tag{5.43}$$

6 Examples of band structure calculations

In this section we will use the Slater-Koster rules to calculate the band structure of different materials. We begin by considering $Pb_{1-x}Sn_xSe$, which is a material that has a rock salt crystal structure and turns out to be a topological crystalline insulator for some values of x, and thus is of interest in this area. Then we move on to a hypothetical two-component material that has a tetragonal crystal structure.

6.1 Band structure of a real material

In this section we will give an example of how the Slater-Koster rules are used to calculate the bulk band structure in a real material. Since, $Pb_{1-x}Sn_xSe$ is a topological crystalline insulator (see e.g. [29]), we choose to do the calculations for this material.

To do this, we need to find the Hamiltonian matrix and diagonalize it at suitable $\mathbf{k}$-values. Before we do this, however, we give a short description of the material.

The material crystallizes in the rock-salt crystal structure, see Figure 4, where each unit cell consists of one Te-atom and one other atom which is either Pb or Sn. The fraction of unit cells containing an Sn-atom is x.

It is always difficult to describe properties of complicated alloys, so we will make use of the *virtual crystal approximation*. This is briefly described in [30, sec. 3.7.3]. In this approximation, the alloy AB_xC_{1-x}, which is disordered, is replaced by the the alloy AD, where D is a pseudoatom that has properties that are weighted averages over the properties of the B and C atoms, for example the mass of D is given by

$$m_D = xm_B + (1-x)m_C. \tag{6.1}$$

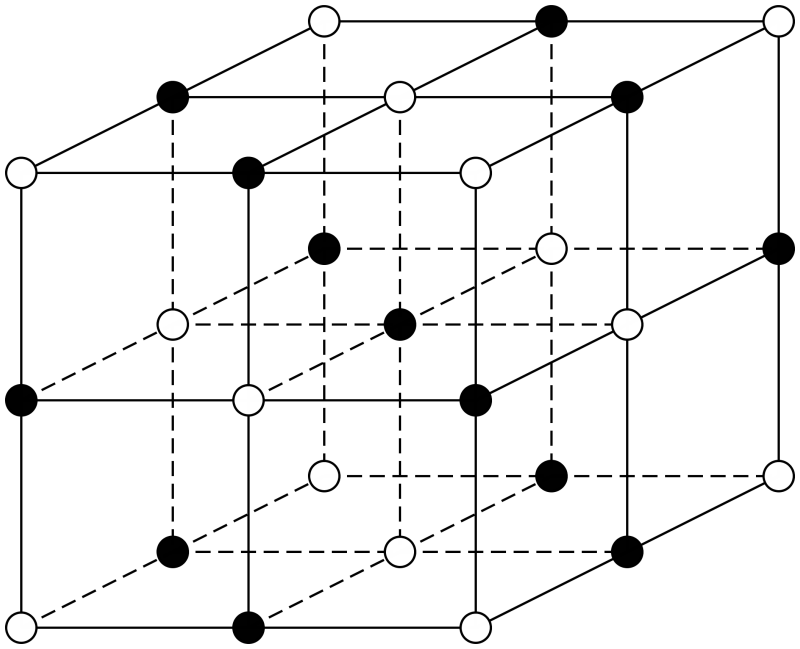

Figure 4: The rocksalt crystal structure

In our case, this would correspond to approximating the properties of $Pb_{1-x}Sn_xSe$ by weighted averages of PbTe and SnTe. These two materials are described in [31] and measured tight-binding parameters are given for them.

6.1.1 Finding the Hamiltonian matrix

The structure of Hamiltonian matrix for the bulk PbTe and SnTe is explicitly given in [31], but we will use the results from [9] to show how we can find it.

The Hamiltonian for the bulk system is given in [31], and has the following form:

$$\begin{aligned} H_0 = &\sum_{\mathbf{R},\sigma,i} \left[|a,i,\sigma,\mathbf{R}\rangle \, E_{i,a} \, \langle a,i,\sigma,\mathbf{R}| + |c,i,\sigma,\mathbf{R}+\mathbf{d}\rangle E_{i,a} \, \langle c,i,\sigma,\mathbf{R}+\mathbf{d}| \right] \\ &+ \sum_{\mathbf{R},\mathbf{R}',\sigma,i,j} \left[|a,i,\sigma,\mathbf{R}\rangle \, V_{i,j} \, \langle c,j,\sigma,\mathbf{R}'+\mathbf{d}| + h.c. \right] \\ &+ \sum_{\mathbf{R},\sigma,\sigma',i} \left[|c,i,\sigma,\mathbf{R}\rangle \, \lambda_c \mathbf{L}_c \bullet \sigma_c \, \langle c,i,\sigma',\mathbf{R}| \right] + \sum_{\mathbf{R},\sigma,\sigma',j} \left[|a,j,\sigma,\mathbf{R}\rangle \, \lambda_a \mathbf{L}_a \bullet \sigma_a \, \langle a,i,\sigma',\mathbf{R}| \right], \end{aligned} \tag{6.2}$$

where the constants E and V correspond to the constants $(\alpha\beta\gamma)$ described in Section 5.1. The last two sums in equation (6.2) correspond to the spin orbit interaction.

In contrast to the discussion about the simple cubic crystal structure described in Section 5.2, we note that this material contains two types of atoms instead of one. This is not a problem, however. The rock-salt structure still greatly resembles the simple cubic structure, and the matrix elements

will therefore be similar to those in Table 2, the difference being that we have to distinguish between the constants that arise from interaction between two atoms of the same type and constants that arise from interaction between atoms of different types. In particular, the structure consists of anions and cations, which we denote by a and c respectively. So instead of one $(sp\sigma)$-integral, we need to distinguish between the cases $(s_ap_a\sigma)$, $(s_ap_c\sigma)$, $(s_cp_a\sigma)$ and $(s_cp_c\sigma)$. We will denote these different cases by e.g. $(\alpha\beta\gamma)^{ac}$.

To approximate the Hamiltonian matrix, in [31] they use the orbitals sp^3d^5 and take spin into account, which means in total 18 orbitals per atom. Since the unit cell of the material contains two atoms, this means that the Hamiltonian matrix, which we now denote by H, becomes a 36×36-matrix. As we know from Section 5.1, the spin is not taken into account in the Slater-Koster theory, but since the spin-orbit interation is only relevant for the self-interaction in certain special cases of the materials, this does not pose any additional difficulties in our derivation.

Now, to find the Hamiltonian matrix, we use the set of orbitals described above as a basis. We label each orbital according to $|$atomic orbital, type of atom, spin$\rangle$, and we get the following set:

$$\begin{array}{llll}
|s,c,\uparrow\rangle & |s,c,\downarrow\rangle & |p_x,c,\uparrow\rangle & |p_y,c,\uparrow\rangle \\
|p_z,c,\uparrow\rangle & |p_x,c,\downarrow\rangle & |p_y,c,\downarrow\rangle & |p_z,c,\downarrow\rangle \\
|d_1,c,\uparrow\rangle & |d_2,c,\uparrow\rangle & |d_3,c,\uparrow\rangle & |d_4,c,\uparrow\rangle \\
|d_5,c,\uparrow\rangle & |d_1,c,\downarrow\rangle & |d_2,c,\downarrow\rangle & |d_3,c,\downarrow\rangle \\
|d_4,c,\downarrow\rangle & |d_5,c,\downarrow\rangle & |s,a,\uparrow\rangle & |s,a,\downarrow\rangle \\
|p_x,a,\uparrow\rangle & |p_y,a,\uparrow\rangle & |p_z,a,\uparrow\rangle & |p_x,a,\downarrow\rangle \\
|p_y,a,\downarrow\rangle & |p_z,a,\downarrow\rangle & |d_1,a,\uparrow\rangle & |d_2,a,\uparrow\rangle \\
|d_3,a,\uparrow\rangle & |d_4,a,\uparrow\rangle & |d_5,a,\uparrow\rangle & |d_1,a,\downarrow\rangle \\
|d_2,a,\downarrow\rangle & |d_3,a,\downarrow\rangle & |d_4,a,\downarrow\rangle & |d_5,a,\downarrow\rangle\,.
\end{array} \tag{6.3}$$

When constructing the Hamiltonian matrix we use the orbitals in this order read from left to right. This means that the Hamiltonian matrix H, will be a 36×36-matrix of the following form:

$$H = \begin{pmatrix} H_{cc} & H_{ca} \\ H_{ac} & H_{aa} \end{pmatrix}, \tag{6.4}$$

where c stands for cation and a stands for anion. These matrices are explicitly given in [31], and we will now show how the non spin dependent parts can be derived using the Slater-Koster rules.

We note that since we only take nearest neighbour interactions into account, a lot of elements in the matrices H_{cc} and H_{aa} are zero. This is because these matrices only contain spin-orbit interactions and interactions between orbitals that belong to the same atom. The methodology for using the Slater-Koster rules to find the Hamiltonian matrix while taking the spin-orbit interaction into account is to simply consider them separately and then adding the results.

We have e.g.

$$\langle p_x,a,\uparrow|H|p_x,c,\uparrow\rangle = (x/x)_1^{ac} = 2(pp\sigma)_1^{ac}\cos\xi + 2(pp\pi)_1^{ac}(\cos\eta + \cos\zeta), \tag{6.5}$$

and

$$\begin{aligned}
\langle d_2,a,\uparrow|H|d_1,c,\uparrow\rangle &= (3z^2 - r^2/x^2 - y^2)_1^{ac} = (x^2 - y^2/3z^2 - r^2)_1^{ca} \\
&= \frac{\sqrt{3}}{2}\left[-(dd\sigma)_1^{ca} + (dd\delta)_1^{ca}\right](\cos\xi - \cos\eta),
\end{aligned} \tag{6.6}$$

where we have used Table 2 to get the matrix elements. We have also introduced some additional notation in order to take the two constituents of the material into account. The upper index tells us which atom the respective orbitals belong to, and the lower index tells us which neighbours we take into account.

Thus the matrices are given by

$$H_{ii} = \begin{pmatrix} H_{sisi} & H_{sipi} & H_{sidi} \\ H_{pisi} & H_{pipi} & H_{pidi} \\ H_{disi} & H_{dipi} & H_{didi} \end{pmatrix}, \tag{6.7}$$

where i denotes the type of atom and thus takes the values a or c, and

$$H_{ac} = \begin{pmatrix} H_{sasc} & H_{sapc} & H_{sadc} \\ H_{pasc} & H_{papc} & H_{padc} \\ H_{dasc} & H_{dapc} & H_{dadc}. \end{pmatrix} \tag{6.8}$$

In H_{ii}, we note that e.g. $H_{sipi} = H^{\dagger}_{pisi}$, since H_{ii} must be Hermitian. We have

$$H_{sisi} = E_{si} I_{2\times 2}, \tag{6.9}$$

$$H_{didi} = E_{di} I_{10\times 10}, \tag{6.10}$$

and

$$H_{pipi} = \begin{pmatrix} E_{pc} & -\mathrm{i}\frac{\lambda_i}{2} & & & & -\frac{\lambda_i}{2} \\ \mathrm{i}\frac{\lambda_i}{2} & E_{pc} & & & & -\mathrm{i}\frac{\lambda_i}{2} \\ & & E_{pc} & -\frac{\lambda_i}{2} & \mathrm{i}\frac{\lambda_i}{2} & \\ & & -\frac{\lambda_i}{2} & E_{pc} & \mathrm{i}\frac{\lambda_i}{2} & \\ & & -\mathrm{i}\frac{\lambda_i}{2} & -\mathrm{i}\frac{\lambda_i}{2} & E_{pc} & \\ \frac{\lambda_i}{2} & \mathrm{i}\frac{\lambda_i}{2} & & & & E_{pc} \end{pmatrix}. \tag{6.11}$$

The matrices H_{pisi}, H_{disi} and H_{dipi} are all zero.

For H_{ac}, we have the following:

$$H_{sasc} = g_0 V_{ss} I_{2\times 2}, \tag{6.12}$$

$$H_{pasc} = \begin{pmatrix} -2g_1 V_{sp} & 0 \\ -2g_2 V_{sp} & 0 \\ -2g_3 V_{sp} & 0 \\ 0 & -2g_1 V_{sp} \\ 0 & -2g_2 V_{sp} \\ 0 & -2g_3 V_{sp} \end{pmatrix}, \tag{6.13}$$

$$H_{sapc} = \begin{pmatrix} -2g_1 V_{ps} & -2g_2 V_{ps} & -2g_3 V_{ps} & 0 & 0 & 0 \\ 0 & 0 & 0 & -2g_1 V_{ps} & -2g_2 V_{ps} & -2g_3 V_{ps} \end{pmatrix}, \tag{6.14}$$

$$H_{papc} = \begin{pmatrix} V_{xx} & & & & & \\ & V_{xx} & & & & \\ & & V_{yy} & & & \\ & & & V_{yy} & & \\ & & & & V_{zz} & \\ & & & & & V_{zz} \end{pmatrix}, \tag{6.15}$$

where

$$V_{xx} = 2g_4V_{pp} + 2(g_5 + g_6)V_{pp\pi}, \tag{6.16}$$

$$V_{yy} = 2g_5V_{pp} + 2(g_4 + g_6)V_{pp\pi}, \tag{6.17}$$

and

$$V_{zz} = 2g_6V_{pp} + 2(g_4 + g_5)V_{pp\pi}. \tag{6.18}$$

We also have

$$H_{dadc} = \begin{pmatrix} H_{53} & 0 \\ 0 & H_{53} \end{pmatrix}, \tag{6.19}$$

and

$$H_{dcda} = \begin{pmatrix} H_{35} & 0 \\ 0 & H_{35} \end{pmatrix}, \tag{6.20}$$

where

$$H_{53} = \begin{pmatrix} -\sqrt{3}g_1V_{pd} & \sqrt{3}g_2V_{pd} & 0 \\ g_1V_{pd} & g_2V_{pd} & -2g_3V_{pd} \\ -2g_2V_{pd\pi} & -2g_1V_{pd\pi} & 0 \\ 0 & -2g_3V_{pd\pi} & -2g_2V_{pd\pi} \\ -2g_3V_{pd\pi} & 0 & -2g_1V_{pd\pi} \end{pmatrix}, \tag{6.21}$$

and H_{35} is the transpose of H_{53} with the order of all p and d exchanged.

Finally, we have that both H_{dasc} and H_{sadc} are zero and that H_{dadc} is given by

$$H_{dadc} = \begin{pmatrix} V_1 & V_{12} & & & & & & & & \\ V_{12} & V_2 & & & & & & & & \\ & & V_3 & & & & & & & \\ & & & V_4 & & & & & & \\ & & & & V_5 & & & & & \\ & & & & & V_1 & V_{12} & & & \\ & & & & & V_{12} & V_2 & & & \\ & & & & & & & V_3 & & \\ & & & & & & & & V_4 & \\ & & & & & & & & & V_5 \end{pmatrix}, \tag{6.22}$$

where

$$V_1 = \frac{3}{2}(g_4 + g_5)V_{dd} + \left(2g_6 + \frac{g_4 + g_5}{2}\right)V_{dd\delta}, \tag{6.23}$$

$$V_2 = \frac{3}{2}(g_4 + g_5)V_{dd\delta} + \left(2g_6 + \frac{g_4 + g_5}{2}\right)V_{dd}, \tag{6.24}$$

$$V_3 = 2(g_4 + g_5)V_{dd\pi} + 2g_6V_{dd\delta}, \tag{6.25}$$

$$V_4 = 2(g_5 + g_6)V_{dd\pi} + 2g_4V_{dd\delta}, \tag{6.26}$$

$$V_5 = 2(g_4 + g_6)V_{dd\pi} + 2g_5V_{dd\delta}, \tag{6.27}$$

and

$$V_{12} = \frac{\sqrt{3}}{2}(g_5 - g_4)(V_{dd} - V_{dd\delta}). \tag{6.28}$$

In the formulas above, we have used the following abbreviations:

$$
\begin{aligned}
g_0(\mathbf{k}) &= 2\left[\cos\left(\frac{k_x a_L}{2}\right) + \cos\left(\frac{k_y a_L}{2}\right) + \cos\left(\frac{k_z a_L}{2}\right)\right], \\
g_1(\mathbf{k}) &= \mathrm{i}\sin\left(\frac{k_x a_L}{2}\right), \\
g_2(\mathbf{k}) &= \mathrm{i}\sin\left(\frac{k_y a_L}{2}\right), \\
g_3(\mathbf{k}) &= \mathrm{i}\sin\left(\frac{k_z a_L}{2}\right), \\
g_4(\mathbf{k}) &= \cos\left(\frac{k_x a_L}{2}\right), \\
g_5(\mathbf{k}) &= \cos\left(\frac{k_y a_L}{2}\right), \\
g_6(\mathbf{k}) &= \cos\left(\frac{k_z a_L}{2}\right),
\end{aligned}
\tag{6.29}
$$

It is clear that for the matrix H to be Hermitian, we must have $H_{ca} = H_{ac}^{\dagger}$, and that H_{aa} and H_{cc} are Hermitian.

The structure of the matrix H is not element-specific, but the parameters are. These can be found in Table 3, where we also write the constants in Slater-Koster notation. Note that some of the parameters are set to zero. This is not a result from fitting the values to experimental data, but rather a choice one makes to ignore the interactions that are very small. [31]

To get the parameters for $\mathrm{Pb}_{1-x}\mathrm{Sn}_x\mathrm{Te}$, we use the virtual crystal approximation, so for an arbitrary parameter P from Table 3, we have

$$
P_{\mathrm{Pb}_{1-x}\mathrm{Sn}_x\mathrm{Te}} = (1-x)P_{\mathrm{PbTe}} + xP_{\mathrm{SnTe}}. \tag{6.30}
$$

We now implement all this information in MATLAB. The code can be found in Section A We diagonalize the matrix along different lines in the reciprocal space. The result for different values of x is shown in Figure 5. We see that the material has a bandgap and therefore is an insulator.

	S-K	PbTe	SnTe
E_{sc}	s_0^{cc}	-7.612	-6.578
E_{sa}	s_0^{aa}	-11.002	-12.067
E_{pc}	p_0^{cc}	3.195	1.659
E_{pa}	p_0^{aa}	-0.237	-0.167
E_{dc}	d_0^{cc}	7.73	8.38
E_{da}	d_0^{aa}	7.73	7.73
λ_c	$-$	1.500	0.592
λ_a	$-$	0.428	0.564
V_{ss}	$(ss\sigma)_1^{ca}$	-0.474	-0.510
V_{sp}	$(sp\sigma)_1^{ca}$	0.705	0.949
V_{ps}	$(sp\sigma)_1^{ac}$	0.633	-0.198
V_{sd}	$(sd\sigma)_1^{ca}$	0	0
V_{ds}	$(sd\sigma)_1^{ac}$	0	0
V_{pp}	$(pp\sigma)_1^{ca}$	2.066	2.218
$V_{pp\pi}$	$(pp\pi)_1^{ca}$	-0.430	-0.446
V_{pd}	$(pd\sigma)_1^{ca}$	-1.29	-1.11
$V_{pd\pi}$	$(pd\pi)_1^{ca}$	0.835	0.624
V_{dp}	$(pd\sigma)_1^{ac}$	-1.59	-1.67
$V_{dp\pi}$	$(pd\pi)_1^{ac}$	0.531	0.766
V_{dd}	$(dd\sigma)_1^{ca}$	-1.35	-1.72
$V_{dd\pi}$	$(dd\pi)_1^{ca}$	0	0
$V_{dd\delta}$	$(dd\delta)_1^{ca}$	0.668	0.618

Table 3: Tight-binding parameters for the alloys PbTe and SnTe.

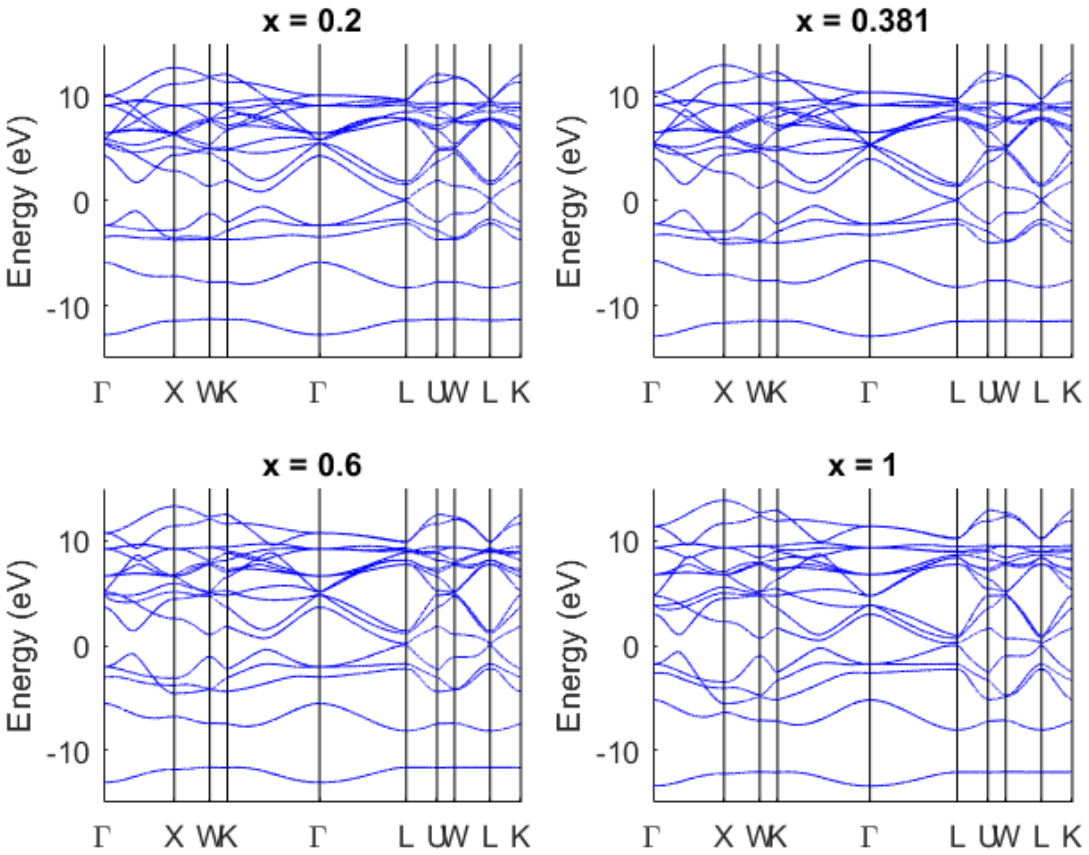

Figure 5: Bulk band structure for $\mathrm{Pb}_{1-x}\mathrm{Sn}_x\mathrm{Te}$ *for* $x = 0.2, 0.381, 0.6, 1$.

Now, we have done calculations for the bulk band structure. Since we are interested in the topological properties of the material, we are actually more interested in the surface states, which one cannot see in the bulk band structure. To see the topological states one would actually have to take the surface into account in the tight-binding calculations. We will show in detail how to do this in the next section for a simpler system. In the case of $Pb_{1-x}Sn_xTe$ we merely show the result obatined in [29], where they have done tight-binding calculations for a slab with 280 atomic layers. This is shown in Figure 6, where we clearly can see when we get gapless surface states. At $x = 0.381$, the gap closes, and for larger x the gap opens again, but gapless surface states remain. This means that for $x \geqslant 0.381$ we get a topological insulator, while for smaller x we get a trivial insulator.

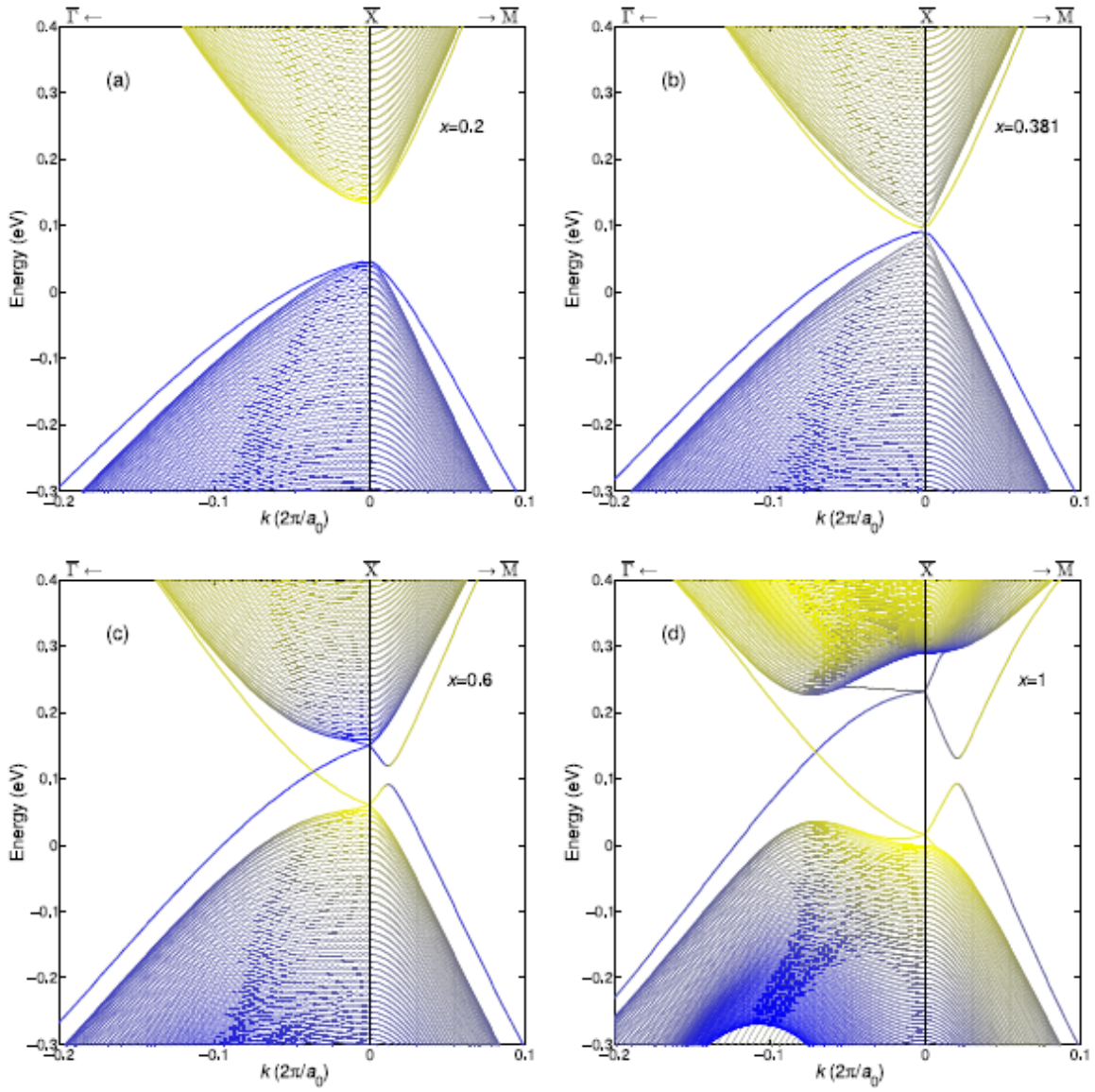

Figure 6: Tight-binding calculations done for a slab of $Pb_{1-x}Sn_xTe$ *with* 280 *layers. Image reprinted with permission from [29].*

6.2 Band structure for a tetragonal lattice

We have previously, in Section 4 described the topological properties of a topological crystaline insulator with C_4 symmetry. Here we will calculate the bandstructure of such a material to demonstrate what such states might look like.

In [8], a tight-binding model of a tetragonal lattice with a unit cell consisting of two inequivalent atoms is described. We will here confirm the results obtained in this simulation and study the dependence of the result on slab thickness. This is not discussed in [8], neither is the lattice constant mentioned or the significance of the numerical values they used.

A unit cell of the tetragonal lattice with two inequivalent basis atoms that is considered is shown in Figure 7. We label the atoms by A and B respectively, and use a coordinate system according to the picture. We are interested in studying states on the (001)-surface, since this is a surface that has the proper symmetry. We model the material as a stack of bilayers consisting of one layer of A-atoms and one layer of B-atoms. In the bulk case, this stack is infinite.

In this model we consider only the p_x- and p_y-orbitals and construct the following bulk Hamiltonian

$$H_{\text{bulk}} = \sum_n \left(H_n^A + H_n^B + H_n^{AB} \right), \tag{6.31}$$

where H^A and H^B represents the intralayer hopping and H^{AB} represents the interlayer hopping. Now, let $\mathbf{r}_i = (x_i, y_i)$ label a position in the xy-plane, $a = A, B$ label the type of atomic layer, α and β label the type of orbital, and let

$$\mathbf{e}_{ij} = \frac{\mathbf{r}_i - \mathbf{r}_j}{|\mathbf{r}_i - \mathbf{r}_j|}. \tag{6.32}$$

Then H^a and H^{AB} can be expressed in the following way:

$$H_n^a = \sum_{i,j} t^a(\mathbf{r}_i - \mathbf{r}_j) \sum_{\alpha\beta} c^{a\alpha\dagger}(\mathbf{r}_i, n) e_{ij}^{\alpha} e_{ij}^{\beta} c_{a\beta}(\mathbf{r}_j, n), \tag{6.33}$$

$$\begin{aligned} H_n^{AB} = &\sum_{ij} t'(\mathbf{r}_i - \mathbf{r}_j) \left[\sum_{\alpha} c_{A\alpha}^{\dagger}(\mathbf{r}_i, n) c_{B\alpha}(\mathbf{r}_j, n) + H.C. \right] \\ &+ t_z' \sum_i \sum_{\alpha} \left[c_{A\alpha}^{\dagger}(\mathbf{r}_i, n) c_{B\alpha}(\mathbf{r}_i, n+1) + H.C \right]. \end{aligned} \tag{6.34}$$

Here $c^\dagger$ and c are creation and annihilation operators, respectively. This means that the terms in the expression are to be interpreted in the following way:

The term

$$\sum_{ij} t'(\mathbf{r}_i - \mathbf{r}_j) c_{A\alpha}^{\dagger}(\mathbf{r}_i, n) c_{B\alpha}(\mathbf{r}_j, n) \tag{6.35}$$

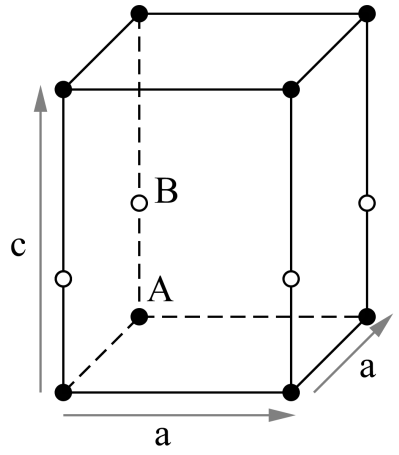

Figure 7: Image of the tetragonal unit cell with lattice constants a and c consisting of two different atoms A and B.

means that we have hopping from site $\mathbf{r}_j$ in the nth B-layer to site $\mathbf{r}_i$ in the nth A-layer. The term

$$t'_z \sum_i \sum_\alpha \left[c^\dagger_{A\alpha}(\mathbf{r}_i, n) c_{B\alpha}(\mathbf{r}_i, n+1) \right], \tag{6.36}$$

means that we have hopping from site $\mathbf{r}_j$ in the $n+1$st B-layer to site $\mathbf{r}_i$ in the nth A-layer. The term

$$\sum_{i,j} t^a(\mathbf{r}_i - \mathbf{r}_j) c^{a\alpha\dagger}(\mathbf{r}_i, n) e^\alpha_{ij} e^\beta_{ij} c_{a\beta}(\mathbf{r}_j, n) \tag{6.37}$$

means we have hopping from site $\mathbf{r}_j$ in the nth a-layer to site $\mathbf{r}_i$ in the nth a-layer. The scale-factor $e^\alpha_{ij} e^\beta_{ij}$ takes into account that the intralayer hopping is of σ-bonding type, which means that it depends on the relative orientation between the p-orbital and the hopping direction $\mathbf{e}^{ij}$.

6.2.1 Hamiltonian matrix for the bulk

In our tight-binding model we take nearest and next-nearest neighbour interactions into account. We denote the intralayer hopping amplitudes by t^a_1 and t^a_2 respectively, while the interlayer hopping amplitudes are denoted by t'_1 and t'_2. This gives us the following Bloch Hamiltonian in matrix form:

$$H(\mathbf{k}) = \begin{pmatrix} H^A(\mathbf{k}) & H^{AB}(\mathbf{k}) \\ H^{AB\dagger}(\mathbf{k}) & H^B(\mathbf{k}) \end{pmatrix}, \tag{6.38}$$

where

$$H^a(\mathbf{k}) = 2t^a_1 \begin{pmatrix} \cos(k_x) & 0 \\ 0 & \cos(k_y) \end{pmatrix} + 2t^a_2 \begin{pmatrix} \cos(k_x)\cos(k_y) & \sin(k_x)\sin(k_y) \\ \sin(k_x)\sin(k_y) & \cos(k_x)\cos(k_y) \end{pmatrix}, \tag{6.39}$$

and

$$H^{AB}(\mathbf{k}) = \left[t'_1 + t'_2(\cos(k_x) + \cos(k_y)) + t'_z e^{\mathrm{i}k_z} \right] I. \tag{6.40}$$

These matrices are determined in [8] by taking the Fourier transform of equations (6.33) and (6.34). We will here instead show how they can immediately be determined from the Slater-Koster rules described in Section 5. The matrices H^A and H^B describe the intralayer interactions. We can therefore model this as a two-dimensional square lattice. From Table 2 we can hence directly read off what the elements in the matrices should be, taking into account that we ignore all terms that contain a k_z-compontent, since we are dealing with a two-dimensional system. We have only p_x- and p_y-orbitals. We have

$$(x/x) = 2(pp\sigma)_1 \cos(ak_x) + 2(pp\pi)_1 \cos(ak_y) + 2(pp\sigma)_2 \left[\cos(ak_x)\cos(ak_y)\right] \tag{6.41}$$

and

$$(x/y) = -2\left[(pp\sigma)_2 - (pp\pi)_2\right] \sin(ak_x)\sin(ak_y). \tag{6.42}$$

Now, according to [8] the intralayer interaction is of σ-bonding type. Thus, we can set the $(pp\pi)$-integrals to 0 in the previous expressions. This gives us

$$(x/x) = 2(pp\sigma)_1 \cos(ak_x) + +2(pp\sigma)_2(\cos(ak_x)\cos(ak_y)) \tag{6.43}$$

and

$$(x/y) = -2(pp\sigma)_2 \sin(ak_x)\sin(ak_y). \tag{6.44}$$

Letting $t_1^a = (pp\sigma)_1$ and $t_2^a = (pp\sigma)_2$, we see that we get back the matrix in equation (6.39).

Now we come to the matrix H^{AB} which describes the interlayer interactions. Its form is not as obvious, since the matrix elements cannot be immediately read off the table. Instead, we note the following: Each A-atom interacts with the B-atoms in the same AB-layer and the B-atoms in the neighbouring AB-layer. Since we are only taking nearest and next-nearest neighbour interactions into account, it is clear that we have to limit ourselves to which of these AB-interactions we are looking at. We choose those in Figure 8 and 9. Now, for the intra-bilayer-interctions, we can again use the results from a square lattice in Table 2. In addition we have to take the interlayer interaction into account which accounts for the periodicity in the z-direction of the material. We approximate the A-atom to interact only with B-atoms in the bilayer closest to A. This intercation takes into account the periodicity of the material in the z-direction, and thus we get an exponential. Thus, for each type of orbital (i.e. p_x and p_y), we get terms

$$t_1' + t_2' \left[\cos(k_x) + \cos(k_y)\right] + t_z' e^{\mathrm{i}k_z}, \tag{6.45}$$

where t_1' is the hopping between A and B atoms in the same bilayer that have the same position in the xy-plane, t_2' is the hopping between A and B atoms in the same bilayer at different positions in the xy-plane. When we take the Hermitian conjugate of H^{AB} we see that the exponent in the exponential term changes sign, which is reasonable, since this also accounts for the interaction between the bilayers in the other direction.

Using the parameters in Table 4, which are taken from [8], and diagonalizing the matrix $H(\mathbf{k})$ for different values of $\mathbf{k}$, we get the band structure seen in Figure 10. This clearly shows that we have a bandgap in the material.

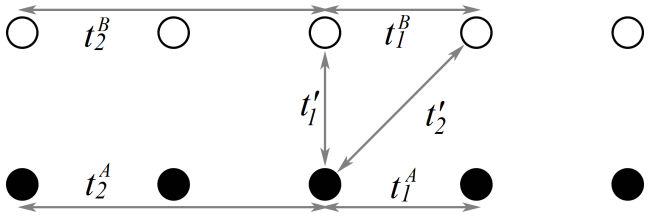

Figure 8: Interactions within each bilayer of the tetragonal lattice.

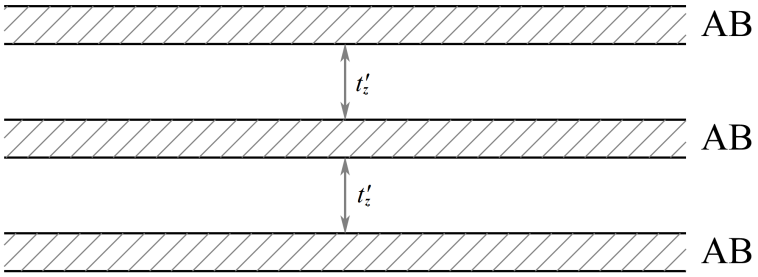

Figure 9: Interactions between the bilayers in the tetragonal structure.

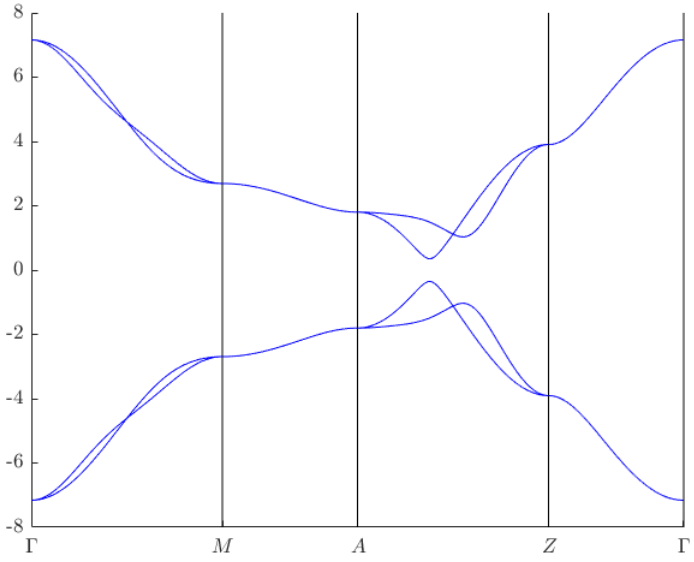

Figure 10: Band structure of a material with tetragonal lattice, calculated using p_x and p_y orbitals and the parameters in Table 4.

6.2.2 Hamiltonian matrix for a slab

In the previous section we found the Hamiltonian matrix for a bulk tetragonal lattice using the p_x and p_y orbitals. Now we want to do band calculations in order to study the surface states of the material. We will do this using the so called *slab method*.

Basically, we first find a matrix describing the two-dimensional structure of an AB-bilayer. We see from the previous sections, that the only ingredient of the matrix H which is *not* part of the bilayer interactions, is the term containing k_z. This means that the bilayer interaction is described by the matrix

$$H(\mathbf{k}) = \begin{pmatrix} H^A(\mathbf{k}) & H^{AB}(\mathbf{k}) \\ H^{AB\dagger}(\mathbf{k}) & H^B(\mathbf{k}) \end{pmatrix}, \tag{6.46}$$

where

$$H^a(\mathbf{k}) = 2t_1^a \begin{pmatrix} \cos(k_x) & 0 \\ 0 & \cos(k_y) \end{pmatrix} + 2t_2^a \begin{pmatrix} \cos(k_x)\cos(k_y) & \sin(k_x)\sin(k_y) \\ \sin(k_x)\sin(k_y) & \cos(k_x)\cos(k_y) \end{pmatrix} \tag{6.47}$$

and

$$H^{AB}(\mathbf{k}) = \left[t_1' + t_2'(\cos(k_x) + \cos(k_y))\right] I. \tag{6.48}$$

Now, suppose we have a slab consisting of N such bilayers on top of each other. We need to connect those to one another through some interaction. This interaction is the same as in the bulk case, with the exception that we do not have periodicity, so instead of the term $t_z' \exp(\mathrm{i}k_z)$, we only get t_z'. The matrix that connects the bilayers is thus of the form

$$H_I = \begin{pmatrix} 0 & 0 \\ H_i & 0 \end{pmatrix}, \tag{6.49}$$

where

$$H_i = \begin{pmatrix} t_z' & 0 \\ 0 & t_z' \end{pmatrix}. \tag{6.50}$$

This gives us the following Hamiltonian matrix for a slab:

$$H_{\text{slab}} = \begin{pmatrix} H & H_I & & \\ H_I^\dagger & \ddots & \ddots & \\ & \ddots & & H_I \\ & & H_I^\dagger & H \end{pmatrix}, \tag{6.51}$$

which we also can write as

$$H_{\text{slab}} = \begin{pmatrix} H^A & H^{AB} & 0 & \dots & \\ H^{AB\dagger} & H^B & H_i & 0 & \dots \\ 0 & H_i & H^A & H_1^{AB} & \ddots \\ 0 & 0 & H_1^{AB\dagger} & H^B & H_i \\ & & & \ddots & \end{pmatrix}. \tag{6.52}$$

The matrix H_{slab} is a $4N \times 4N$ matrix, where N is the number of layers in the slab.

Table 4: Parameters used in the tight-binding calculation for a material with tetragonal lattice.

t_1^A	1
t_2^A	0.5
t_1^B	-1
t_2^B	-0.5
t_1'	2.5
t_2'	0.5
t_z'	2

In order to find the band structure of the slab, we need to diagonalize the H_{slab}. The result will depend on how many layers we use in the slab, but the band structure is expected to converge for large enough N. We will examine the properties of such matrices in more detail in Section 7. For some different numbers of layers, we get the structures shown in Figure 11. We see that the band structure appears to converge quickly. Using 20 layers we get the result shown in Figure 12, and comparing with the plots in Figure 11, it is reasonable to expect that this is a good approximation of the band structure. At this point we see that we get metallic surface states at the $\bar{M}$-point which have parabolic dispersion.

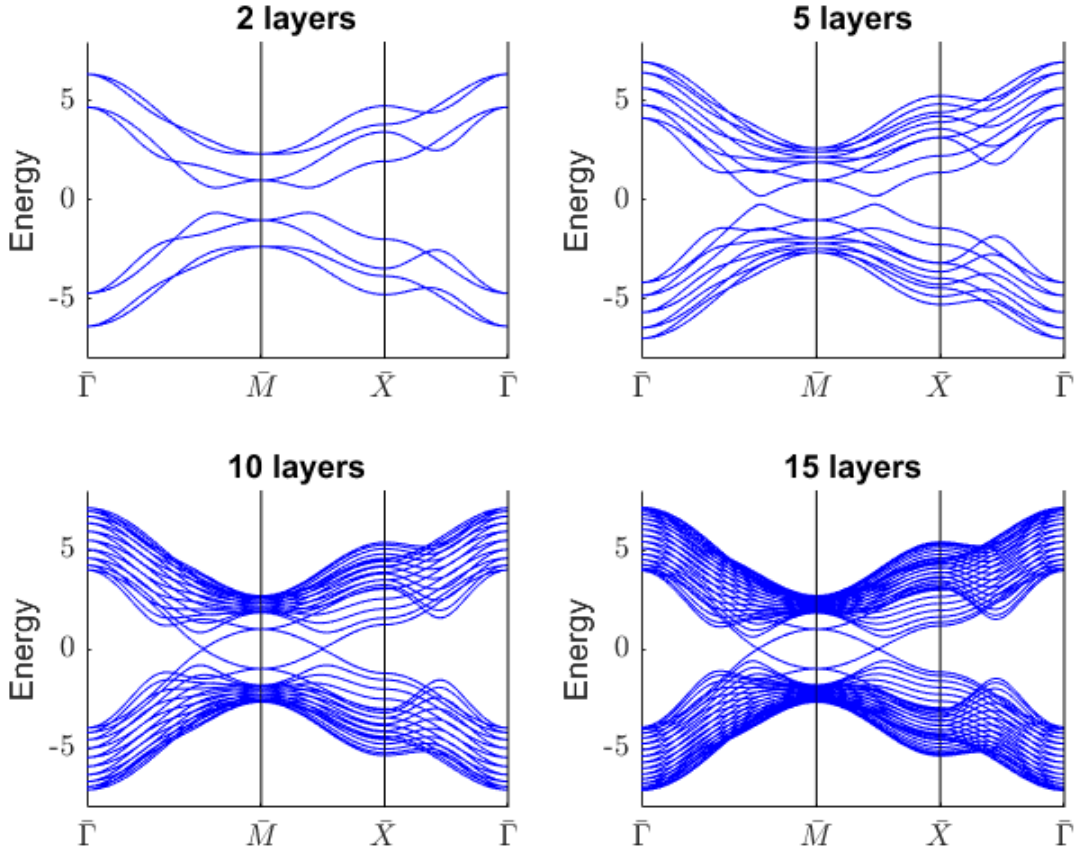

Figure 11: Band structure for slabs with different number of layers.

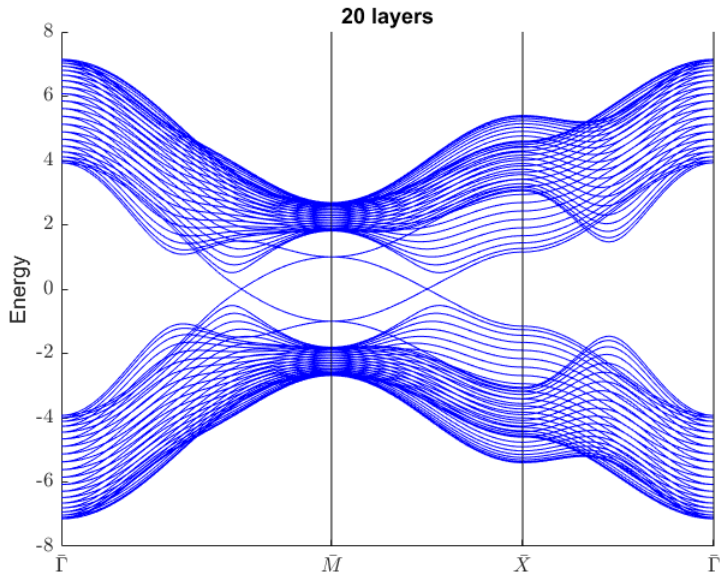

Figure 12: Band structure for a slab with 20 layers.

7 Eigenvalues of block-tridiagonal matrices

As we have seen in the previous sections, the matrices that arise in tight-binding calculations are on block-tridiagonal form, i.e. we have matrices of the form

$$M = \begin{pmatrix} A & B & & & \\ B^\dagger & A & B & & \\ & B^\dagger & \ddots & \ddots & \\ & & \ddots & \ddots & B \\ & & & B^\dagger & A \end{pmatrix}, \tag{7.1}$$

where A is a square Hermitian matrix and B is a square matrix which is not necessarily invertible or Hermitian. It is thus of interest to study the behaviour of the eigenvalues of such matrices, in order to be able to predict important features of the result. In particular, we want to describe the eigenvalues using properties of the matrices A and B and study how the distribution of eigenvalues is affected when the size of M increases.

We will denote the eigenvalues of an $n \times n$ matrix X by $\lambda_i(X)$, ordered according to

$$\lambda_1(X) \geqslant \lambda_2(X) \geqslant \cdots \geqslant \lambda_n(X). \tag{7.2}$$

We let A and B be $n \times n$-matrices and assume that we have N A:s on the diagonal of M, i.e. M is a $Nn \times Nn$-matrix. In cases where we want to stress the size of M, we write M_N. Now, we see that we can write

$$M = \tilde{A} + \tilde{B}, \tag{7.3}$$

where

$$\tilde{A} = \begin{pmatrix} A & & & & \\ & A & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & A \end{pmatrix}, \tag{7.4}$$

and

$$\tilde{B} = \begin{pmatrix} 0 & B & & & \\ B^\dagger & 0 & B & & \\ & B^\dagger & \ddots & \ddots & \\ & & \ddots & \ddots & B \\ & & & B^\dagger & 0 \end{pmatrix}. \tag{7.5}$$

From [32], we have the following result:

Proposition 7.1. *Let X_1 and X_2 be two square matrices of the same size, and denote their respective ith eigenvalues by $\lambda_i(X_1)$ and $\lambda_i(X_2)$. Then*

$$|\lambda_i(X_1) - \lambda_i(X_2)| \leqslant \|X_1 - X_2\|_2 , \tag{7.6}$$

where $\|-\|_2$ denotes the 2-norm of a matrix.

The two-norm of a matrix X is given by

$$\|X\|_2 = \sigma_{\max}(X), \tag{7.7}$$

where $\sigma_{\max}(X)$ denotes the largest singular value of X.

We are interested in a bound on the eigenvalues of M. From (7.6), we have

$$|\lambda_i(M) - \lambda_i(\tilde{A})| \leqslant \left\|M - \tilde{A}\right\|_2 = \left\|\tilde{B}\right\|_2. \tag{7.8}$$

The eigenvalues of $\tilde{A}$ are the same as those for A but with N times the multiplicity, so these are known. This gives us

$$\lambda_i(M) \leqslant \lambda_i(\tilde{A}) \pm \sigma_{\max}(\tilde{B}), \tag{7.9}$$

and we see that the eigenvalues of M will be centered around the eigenvalues of A at a maximum distance of the largest singular value of $\tilde{B}$. So what remains to be done is to calculate the singular values of $\tilde{B}$. First we will present some standard theory about the singular value decomposition, and then we move on to our particular case.

7.1 The singular value decomposition

The singular value decomposition of a matrix is a sort of generalization of diagonalization of quadratic matrices. It is e.g. described in [33].

Theorem 7.2. *Let X be an $m \times n$ matrix with rank r. Then there exists an $m \times n$ matrix*

$$\Sigma = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix}, \tag{7.10}$$

where D is an $r \times r$ diagonal matrix with entries $\sigma_i \geqslant 0$, and an $m \times m$ unitary matrix U and an $n \times n$ unitary matrix V such that

$$X = U\Sigma V^\dagger. \tag{7.11}$$

The diagonal entries σ_i in D are uniquely determined by A and called the *singular values* of A, and any decomposition of a matrix A as described in equation (7.11) is called a *singular value decomposition*.

Singular values are closely related to eigenvalues. Namely, the singular values of a matrix X are the non-negative square roots of the eigenvalues of $X^\dagger X$. We note that for a unitarily diagonalizable matrix X, we have

$$X^\dagger X = (UDU^{-1})^\dagger (UDU^{-1}) = UDU^\dagger UDU^{-1} = UD^2U^{-1}, \tag{7.12}$$

where D is a diagonal matrix and U is unitary. We see that this means that the eigenvalues of $X^\dagger X$ are the eigenvalues of X squared. This in turn means that if X is diagonalizable, the singular values of X are the absolute values of the eigenvalues of X.

7.2 Singular values of $\tilde{B}$

We first present a general result about tridiagonal Toeplitz matrices, i.e. tridiagonal matrices with constant elements along the diagonals. According to e.g. [34] we have the following:

Proposition 7.3. *The $n \times n$ matrix*

$$\begin{pmatrix} a & b & & \\ c & \ddots & \ddots & \\ & \ddots & \ddots & b \\ & & c & a \end{pmatrix} \tag{7.13}$$

has eigenvalues

$$\lambda = a + 2\sqrt{bc}\cos\left(\frac{k\pi}{n+1}\right). \tag{7.14}$$

This result will be useful later. For now, we begin by noting that $\tilde{B}$ in equation (7.5) is a Hermitian matrix. This means that it is normal, which in turn means that the absolute values of the eigenvalues of $\tilde{B}$ are the singular values of $\tilde{B}$. Thus, we want to determine the eigenvalue of $\tilde{B}$ with the largest absolute value. Depending on the properties of $\tilde{B}$, we will get different results. In particular, the results will differ depending on if B is diagonilizable or not. For the diagonalizable case, we follow the reasoning in [35] when they find eigenvalues for the discretization of the Poisson equation, and arrive at the following result:

Proposition 7.4. *Let B be a diagonalizable matrix. The eigenvalues of $\tilde{B}$ are given by*

$$\lambda_{j,k} = 2|\lambda_j(B)|\cos\left(\frac{k\pi}{m+1}\right), \tag{7.15}$$

where $\lambda_j(B)$ are the eigenvalues of B.

Proof. Let λ be an eigenvalue of B and let u be the corresponding eigenvector. We partition this eigenvector in the following way:

$$u = \begin{pmatrix} u_1 \\ \vdots \\ u_m \end{pmatrix}, \tag{7.16}$$

where

$$u_l = \begin{pmatrix} u_{1,l} \\ \vdots \\ u_{n,l} \end{pmatrix}, \quad l = 1, \ldots, m. \tag{7.17}$$

This means that the equation

$$\tilde{B}u = \lambda u, \tag{7.18}$$

can be rewritten as

$$B^\dagger u_{l-1} - \lambda I u_l + B u_{l+1} = 0, \quad l = 1, \ldots, m, \tag{7.19}$$

where we have set $u_0 = u_{m+1} = 0$. Now, $\tilde{B}$ is unitarily diagonalizable, and we can write it as

$$\tilde{B} = UDU^\dagger, \tag{7.20}$$

where U is a unitary matrix, and D is diagonal with diagonal elements $\lambda_j(B)$. We can now rewrite equation (7.19) in the following way:

$$UD^\dagger U^\dagger u_{l-1} - \lambda UU^\dagger u_l + UDU^\dagger u_{l+1} = 0. \tag{7.21}$$

Multiplying this from the left by $U^\dagger$ and setting $y_l = U^\dagger u_l$ gives us

$$D^\dagger y_{l-1} - \lambda y_l + Dy_{l+1} = 0. \tag{7.22}$$

Since the only matrices we have in this equation are diagonal, we can write the equation element-wise instead:

$$\lambda_j^*(B) y_{j,l-1} + \lambda_j(B) y_{j,l+1} = \lambda y_{j,l}. \tag{7.23}$$

Now we consider a fixed value of j. We see that equation (7.23) is satisfied if the vector $(y_{j,1}, \ldots, y_{j,m})$ is an eigenvector with eigenvalue λ of the matrix

$$\begin{pmatrix} 0 & \lambda_j(B) & & \\ \lambda_j^*(B) & \ddots & \ddots & \\ & \ddots & \ddots & \lambda_j(B) \\ & & \lambda_j^*(B) & 0 \end{pmatrix}. \tag{7.24}$$

The eigenvalues of this matrix are, according to Proposition 7.3

$$\lambda = 2|\lambda_j(B)| \cos\left(\frac{k\pi}{m+1}\right), \tag{7.25}$$

which proves the proposition. □

Now we want to use this result to estimate the eigenvalues of M. We have the following result:

Proposition 7.5. *Let B be a diagonalizable matrix. Then the eigenvalues of M are in the interval*

$$\lambda_i(M) \leqslant \lambda_i(\tilde{A}) \pm 2\max|\lambda_j(B)|. \tag{7.26}$$

Proof. From equation (7.9) we see that we need to find the largest singular value of $\tilde{B}$, which in this case is the eigenvalue of $\tilde{B}$ with the largest absolute value. From Proposition 7.4, we see this singular value satisfies

$$\sigma_{\max}(\tilde{B}) \leqslant 2\max|\lambda_j(B)|. \tag{7.27}$$

Inserting this into equation (7.9), we get

$$\lambda_i(M) \leqslant \lambda_i(\tilde{A}) \pm 2|\lambda_j(B)|, \tag{7.28}$$

which proves the statement. □

This means that the eigenvalues of M are gathered in bands of maximum width two times the maximum eigenvalue of B centered around the eigenvalues of A.

Often, however, the matrix B is of the form

$$B = \begin{pmatrix} 0 & 0 \\ C & 0 \end{pmatrix}, \tag{7.29}$$

which means that it is not diagonalizable (since $B^2 = 0$). In this case the proof of Proposition 7.4 is not valid, and we have to change the reasoning.

We get the following result:

Proposition 7.6. *Let $B = \begin{pmatrix} 0 & 0 \\ C & 0 \end{pmatrix}$. Then the singular values of $\tilde{B}$ are given by*

$$\sigma(\tilde{B}) = \sqrt{\lambda(CC^\dagger)}. \tag{7.30}$$

Proof. We begin by noting that

$$\tilde{B}\tilde{B}^\dagger = \begin{pmatrix} B^2 & 0 & BB^\dagger & & & \\ 0 & (B^\dagger)^2 + B^2 & \ddots & \ddots & & \\ B^\dagger B & \ddots & \ddots & \ddots & \ddots & \\ & \ddots & \ddots & \ddots & \ddots & BB^\dagger \\ & & \ddots & \ddots & (B^\dagger)^2 + B^2 & 0 \\ & & & B^\dagger B & 0 & (B^\dagger)^2 \end{pmatrix}. \tag{7.31}$$

Since both B^2 and $(B^\dagger)^2$ are zero, this reduces to

$$\tilde{B}\tilde{B}^\dagger = \begin{pmatrix} 0 & 0 & BB^\dagger & & & \\ 0 & \ddots & \ddots & \ddots & & \\ B^\dagger B & \ddots & \ddots & \ddots & \ddots & \\ & \ddots & \ddots & \ddots & \ddots & BB^\dagger \\ & & \ddots & \ddots & \ddots & 0 \\ & & & B^\dagger B & 0 & 0 \end{pmatrix}. \tag{7.32}$$

Now, it would be nice if one could reuse the argument in Proposition 7.4. However, even though $B^\dagger B$ and $BB^\dagger$ have the same eigenvalues, they do not have the same eigenvectors, which means that we cannot diagonalize them in the same way as we did in the proof of Proposition 7.4, and thus we need some other strategy. We note that if B is of the form (7.29), we have

$$BB^\dagger = \begin{pmatrix} 0 & 0 \\ 0 & CC^\dagger \end{pmatrix}, \tag{7.33}$$

and

$$B^\dagger B = \begin{pmatrix} C^\dagger C & 0 \\ 0 & 0 \end{pmatrix}. \tag{7.34}$$

Inserting this into the expression for $\tilde{B}\tilde{B}^\dagger$, we see that we get a matrix that is similar to

$$\begin{pmatrix} C^\dagger C & & & & \\ & CC^\dagger & & & \\ & & C^\dagger C & & \\ & & & \ddots & \\ & & & & CC^\dagger \end{pmatrix}. \tag{7.35}$$

Thus $\tilde{B}\tilde{B}^\dagger$ has the same eigenvalues as $CC^\dagger$ (but with different multiplicity), which means that the singular values of $\tilde{B}$ are given by

$$\sigma(\tilde{B}) = \sqrt{\lambda(CC^\dagger)}. \tag{7.36}$$

□

As a consequence of this we have

Proposition 7.7. *Let* $B = \begin{pmatrix} 0 & 0 \\ C & 0 \end{pmatrix}$. *Then the eigenvalues of* M *are in the intervals*

$$\lambda_i(M) \leqslant \lambda_i(\tilde{A}) \pm \max\sqrt{\lambda_j(CC^\dagger)}. \tag{7.37}$$

Proof. From Proposition 7.6 we immediately have

$$\sigma_{\max}(\tilde{B}) = \max\sqrt{\lambda_j(CC^\dagger)}, \tag{7.38}$$

which means

$$\lambda_i(M) \leqslant \lambda_i(\tilde{A}) \pm \sqrt{\lambda_{\max}(CC^\dagger)}. \tag{7.39}$$

□

These results explain why the block tridiagonal matrices give rise to the band structures that we have seen in Section 6. Namely, because of the form of the matrix, the eigenvalues will naturally be contained in bands. What this does not explain, however, is where the surface states come from, since these are distinct from the ordinary band structure of the bulk.

In Figures 13, 14 and 15 the eigenvalues of M_N are shown for different choices of A and B as a function of N. The Hermitian matrices A and B are constructed by randomly generating elements in the interval $(0, 1)$, but the elements of A are multiplied by 10 in order for the effect of the eigenvalue distribution to be clear. We see that the eigenvalues are separated into bands. The horizontal lines in Figures 13 and 14 show the eigenvalues of A and $\lambda(A) \pm \sigma_{\max}(B)$. We see that the eigenvalues are contained within the bands described. Now, a relevant question to ask oneself in this case is whether the other eigenvalues of B will matter, or if it is only the largest one that is relevant. This case is shown in Figure 15, where the horizontal lines correspond to $\lambda(A) \pm \lambda B$, for all eigenvalues of B, but there is no obvious relation to the other eigenvalues, especially when also taking the result in Figures 13 and 14 into account.

Now, in all these three plots, we clearly see that the bands are non-overlapping. This, however is not always the case, and to demonstrate this, we plot the eigenvalues of $H_{\text{slab}}(\pi, \pi, 0)$, described in

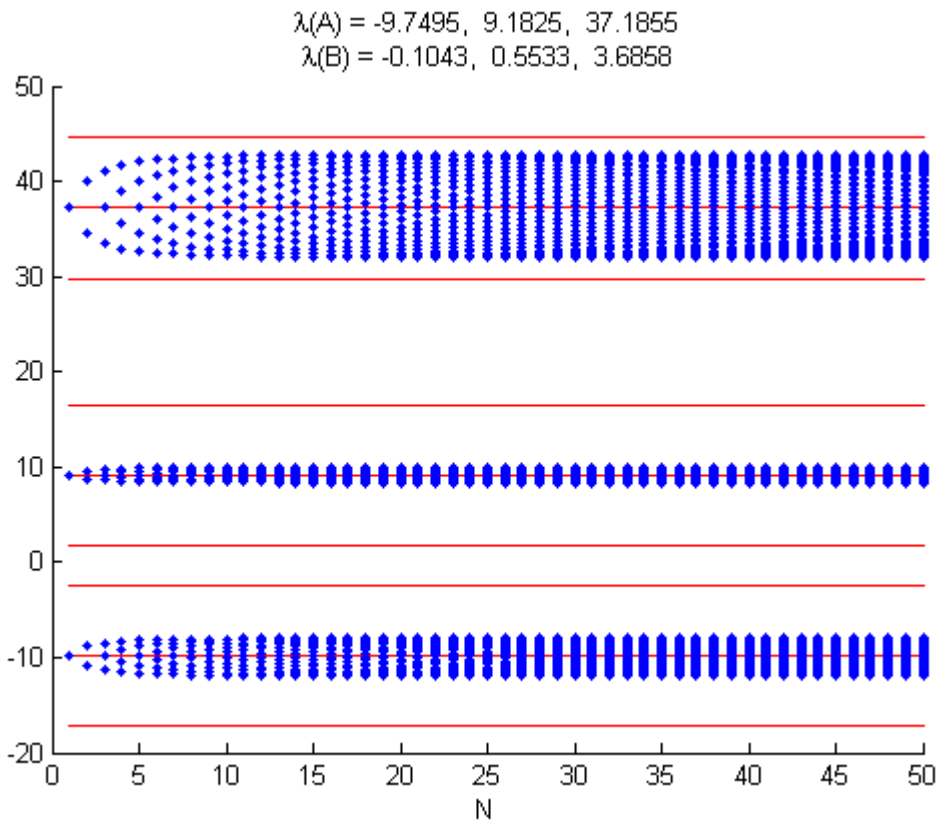

Figure 13: Eigenvalues of the matrix M_N as a function of N. The eigenvalues of A and B are shown in the image.

Section 6.2. This is shown in Figure 16. In this case the matrix B is not diagonalizable, but of the form in Proposition 7.6, so the horizontal lines are determined differently from those in the other figures.

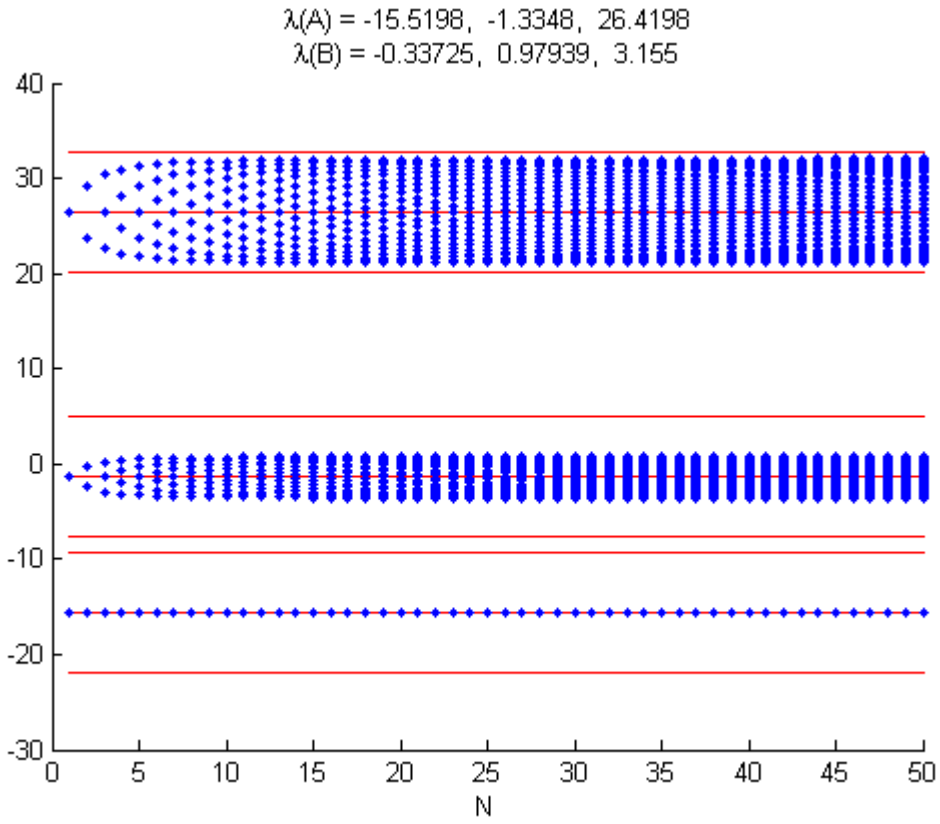

Figure 14: Eigenvalues of the matrix M_N as a function of N. The eigenvalues of A and B are shown in the image.

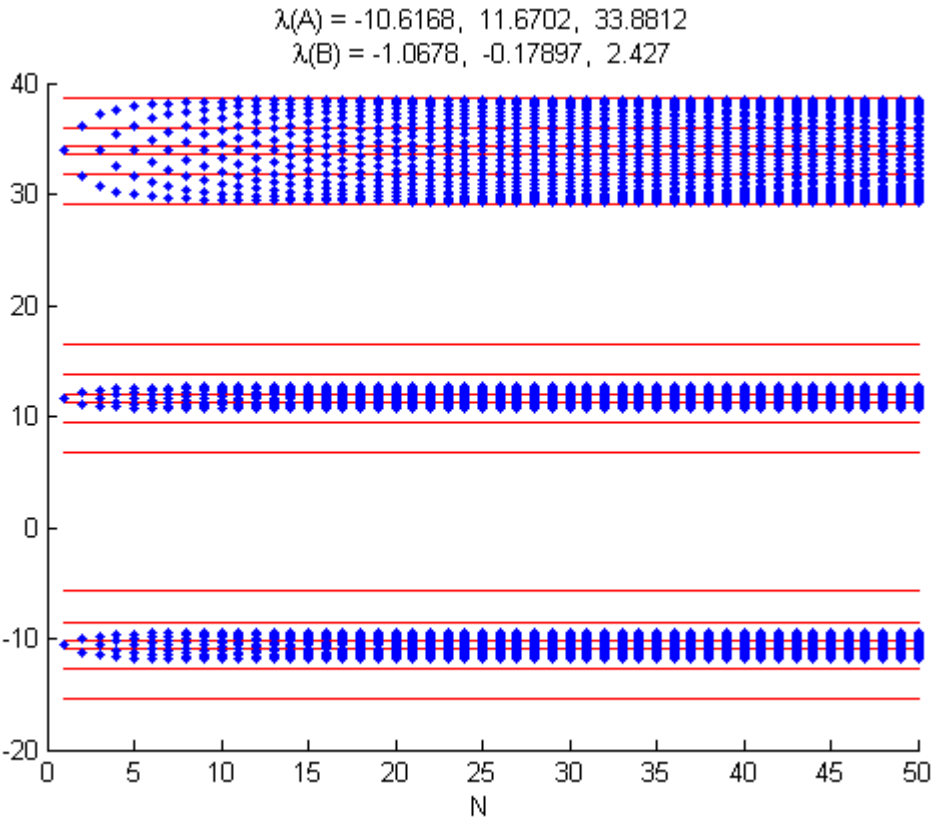

Figure 15: Eigenvalues of the matrix M_N as a function of N. The eigenvalues of A and B are shown in the image.

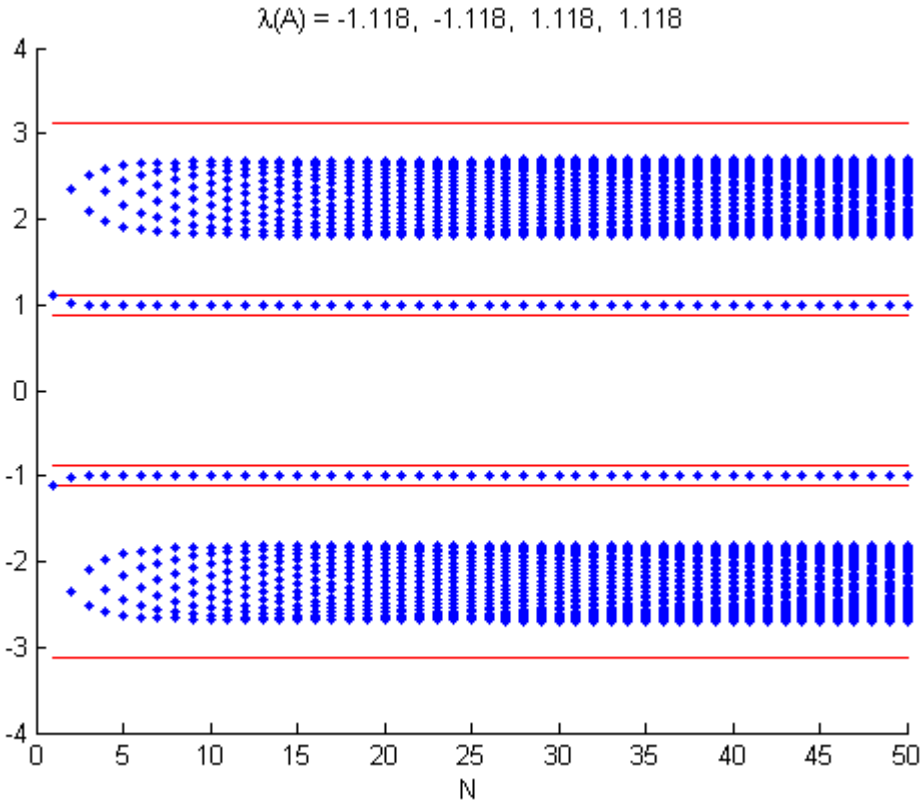

Figure 16: Eigenvalues of the matrix $M_N = H_{\mathrm{slab}}(\pi, \pi, 0)$ as a function of N.

We note that if the elements of A and B are similar in size, the overlap between the bands is bigger. This is related to the fact that if we scale the elements of a matrix by some number, we scale the eigenvalues of the matrix by the same number. So we can always choose a γ, that is large enough such that

$$\min\left[\lambda_{i+1}(\gamma A) - \lambda_i(\gamma A)\right] > \max|\lambda_j(B)|. \tag{7.40}$$

It seems like $\gamma = 10$ is enough in most cases that we consider where the elements of A and B all are in the interval $(0, 1)$.

7.3 Convergence of the eigenvalues

Now, in order to do proper band structure calculations, one needs the distribution of eigenvalues to converge when one increases the slab-thickness. We have previously shown that the eigenvalues are limited to appear in bands that are determined by the eigenvalues of the A and B matrices. Also, Figures 13-16 makes it seem likely that they do converge. This, however, does not really tell us if the correct band structure can be expected to be found after a certain number of iterations. This is what we will examine next.

From [32] we have the following result:

Proposition 7.8. *Consider a Hermitian $n \times n$-matrix X and let Y be a principal submatrix of X of order $n - k$. Denote the eigenvalues of X by $\mu_1 \geqslant \mu_2 \geqslant \cdots \geqslant \mu_n$ and the eigenvalues of Y by $\nu_1 \geqslant \nu_2 \geqslant \cdots \geqslant \nu_{n-k}$. Then we have*

$$\mu_i \geqslant \nu_i \geqslant \mu_{i+k}, \tag{7.41}$$

for $i = 1, \ldots, n - k$.

Now, let M_N denote the matrix M with N repetitions of the matrix A on the diagonal. Then we see that M_{N-1} is a principal submatrix of M_N and we can make use of the result. An immediate consequence of Proposition 7.8 is that $\lambda_1(M_N)$, considered as a function of N, is increasing. Since we also have proven that we have an upper bound on the eigenvalues of M_N, we see that $\lambda_1(M_N)$ forms a convergent sequence. Also, the same must be true for $\lambda_{nN}(M_N)$ (i.e. the smallest eigenvalue of M_N). This is because $-M_N$ has the same eigenvalues as M_N, but with reversed signs, which means that the smallest eigenvalue of M_N is the largest eigenvalue of $-M_N$.

This, however, only says that the largest and smallest eigenvalues converge to some value. In order to make any statements about the band structure as a whole, we need to examine the other eigenvalues as well.

7.3.1 Matrices with non-overlapping eigenvalue bands

For simplicity, we begin by assuming that the bands of eigenvalues do not overlap. In particular, this means that we can rewrite the matrix M in the following way:

$$M_N = UD_NU^\dagger = UD_1^NU^\dagger + UD_2^NU^\dagger + \ldots UD_n^NU^\dagger, \tag{7.42}$$

where n is the size of A and the matrix D_i^N is given by

$$D_i^N = \begin{pmatrix} 0 & & & & & & & & \\ & \ddots & & & & & & & \\ & & 0 & & & & & & \\ & & & \lambda_{in} & & & & & \\ & & & & \ddots & & & & \\ & & & & & \lambda_{in+n-1} & & & \\ & & & & & & 0 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 0 \end{pmatrix} \tag{7.43}$$

In particular this decomposition means that the matrix D_i has the eigenvalues that are in the ith band on the diagonal (including a lot of zeros). So $UD_iU^\dagger$ is a matrix with these eigenvalues. Experiments in MATLAB suggest that $UD_iU^\dagger$ has a structure that is similar to that of M, up to a small perturbation. Namely, it appears to be of the form

$$UD_i^NU^\dagger = \begin{pmatrix} \tilde{S}^N & \tilde{T}^N & & & & & \\ (\tilde{T}^N)^\dagger & S & T^N & & & & \\ & (T^N)^\dagger & \ddots & \ddots & & & \\ & & \ddots & \ddots & T^N & & \\ & & & (T^N)^\dagger & S & \tilde{T}^N & \\ & & & & (\tilde{T}^N)^\dagger & \tilde{S}^N & \end{pmatrix} + E, \tag{7.44}$$

where $\tilde{S}^N$ and $\tilde{T}^N$ have approximately the same elements as S^N and T^N respectively and E is an overall small perturbation.

We note that S^N does not have exactly the same elements as S^{N-1}, which is to be expected since the elements of D_i change slightly as N changes. We do, however, expect some sort of convergence towards an S^∞ matrix. The same goes for the T-matrices. So for large N, the relation between $UD_i^NU^\dagger$ and $UD_i^{N+1}U^\dagger$ is approximately that we add another S^∞ on the diagonal and another T^∞ on the subdiagonal.

This means that an approximate version of Proposition 7.8 should hold. Each time we increase N, we add precisely one non-zero eigenvalue to $UD_i^NU^\dagger$. This means that for the eigenvalues within a band, we have approximately

$$\lambda_i(UD_i^NU^\dagger) \geqslant \lambda_{i+1}(UD_i^{N+1}U^\dagger) \geqslant \lambda_{i+1}(UD_i^NU^\dagger), \tag{7.45}$$

which shows that we have (approximate) convergence of the largest and smallest eigenvalues in each band.

7.4 Practicalities of finding the eigenvalues of block-tridiagonal matrices

One aspect that makes these observations beneficial, is that even though one cannot say anything about the precise size of the eigenvalues, one can immediately determine whether there is any use

in looking for surface state crossings or not. Namely, those surface state crossings can only occur if there is an overlap between the bands that contain the eigenvalues. This might be a good thing if the matrices one deals with are large. Diagonalization is a complex process which scales as N^3, where N is the size of the matrix (although, in the special case of the block-tridiagonal matrices discussed here, it is possible that the computational complexity is smaller). Thus, one wants to avoid making unnecessary eigenvalue calculations.

We see that if one were to make slab-calculations for the system described in Section 6.1 the slab Hamiltonian matrix would be of size $36N \times 36N$, where N is the number of layers used in the slab. This is done in [29], where they use 280 layers. This means a 10080×10080-matrix which has to be diagonalized at a large number of different points in the reciprocal space. If one does not have a huge amount of computational power at hand, this is a very time-consuming process – especially if one does not know where to look for the surface level crossings to begin with. One could then narrow down the search area by considering where it is possible to get overlap between the bands.

7.4.1 Observations related to MATLAB

To find the eigenvalues one can use one of either `eig` or `eigs` in MATLAB. As mentioned previously, diagonalization of the relevant matrices used in band structure calculations for slabs, is extremely time-consuming,it is necessary to evaluate which method is the most beneficial to use.

In MATLAB, the function `eig` calculates all eigenvalues of a matrix, but if the matrix contains complex valued elements, `eig` can only be used if the matrix is a full matrix. The function `eigs`, on the other hand, calculates a pre-specified number of the eigenvalues of a matrix. It can be used for both full and sparse matrices.

The Hamiltonian is a Hermitian banded matrix with many zeros, and thus we have two options. Use `eig` on the full Hamiltonian matrix, or use `eigs` on the sparse Hamiltonian matrix. Since we are interested in the band structure, we need to calculate enough eigenvalues, and in order to find interesting parts of the spectrum, one has to be able to run simulations multiple times. Thus it is reasonable to compare the time needed for `eigs` to find the first n eigenvalues with the time needed for `eig` to find all eigenvalues of the matrix. In our simulations, `eigs` is faster for finding less than 450 eigenvalues.

Additional examinations show that the time `eigs` uses to find a certain number of eigenvalues is only weakly dependent on the size of the matrix, so increasing the number of atomic layers does not necessarily give us an eigenvalue problem that the computer cannot handle.

8 Conclusions

In this thesis, we have reviewed some basic theory about topological insulators and topological crystalline insulators. We have seen the role that time-reversal symmetry alone or in combination with other symmetries plays in these materials, and how one can derive topological invariants for these materials. We have also reviewed how the tight-binding method can be used in order to study the band structure of a material. In particular we make such calculations in order to study the

surface states of a topological crystalline insulator. We find that in this case the calculations give us the expected surface level crossing. We conclude the thesis by discussing the properties of the eigenvalues of block tridiagonal matrices, that arise when doing certain tight-binding calculations. We provide a bound on the eigenvalues, and argue that the distribution of eigenvalues should converge as the size of the matrix increases.

A MATLAB-code

A.1 The $Pb_{1-x}Sn_xTe$ system

The main program `PbSnTe`:

```
clear;
close all;
clc;

q = 1; % A counter which determines which plot is to be used.
for x = [0.2 0.381 0.6 1] % Determines for which values of x the band
    % structure is to be calculated.

aL = 0.63e-9*x + 0.646e-9*(1-x); % Calculates the VCA-lattice constant.

% Lists interesting reciprocal lattice points.
Gamma = [0 0 0];
X = [0 1/2 1/2];
L = [1/2 1/2 1/2];
W = [1/4 3/4 1/2];
U = [1/4 5/8 5/8];
K = [3/8 3/4 3/8];

% Creates a matrix which tells us between which points in the reciprocal
% space the band structure is to be calculated.
path_matrix = [Gamma; X; W; K; Gamma; L; U; W; L; K];

% Defines reciprocal lattice vectors.
b1 = 2*pi/aL*[1 -1 1];
b2 = 2*pi/aL*[1 1 -1];
b3 = 2*pi/aL*[-1 1 1];

% Creates a matrix which is used to get a plot with the correct scaling
% between the reciprocal lattice points.
distance_matrix = [];
for m = 1:size(path_matrix,1)-1
    distance_matrix = [distance_matrix norm(path_matrix(m,:)-...
path_matrix(m+1,:))];
end
distance_matrix = distance_matrix/norm(distance_matrix,inf);

% This loop calculates and plots the band structure. In each loop the band
% structure between two of the chosen reciprocal lattice points is
% calculated.
```

```
s = 0;
xplace = [s];
for m = 1:size(path_matrix,1)-1
    % Evaluates at which points the Hamiltonian matrix should be
    % diagonalized.
    step_size = 0.001;
    P1 = path_matrix(m,:);
    P2 = path_matrix(m+1,:);
    [points,t] = line_between_points(P1',P2',[b1' b2' b3'],step_size);

    % Finds the eigenvalues of the Hamiltonian matrix at the desired
    % k-points. This matrix is defined in the function hamiltonian_matrix.
    energies_Delta = [];
    for j = 1:size(points,1)
        k = points(j,:);
        H = hamiltonian_matrix(k(1),k(2),k(3),x);
        E = eig(H);
        energies_Delta = [energies_Delta E];
    end

    % Plots the calculated band-structure.
    for j = 1:36
        figure(1)
        subplot(2,2,q)
        hold on
        plot(t/t(end)*distance_matrix(m)+s,energies_Delta(j,:),'-')
    end

    s = s+distance_matrix(m);
    xplace = [xplace, s];
    line([s,s],ylim,'Color','black')
end

% Makes the plot look nice.
xlim([0 s])
ylim([-15 15])
title(['x = ', num2str(x)])
set(gca, 'xTick', xplace)
set(gca, 'xTickLabel', {'\Gamma','X','W','K','\Gamma','L','U','W','L','K'})
ylabel('Energy (eV)')

q = q+1;
end
```

The functions used in the main program are the following:

```
function [points,t] = line_between_points(P1,P2,basis_vectors,step_size)
% Calculates equidistant points on a line between P1 and P2 (given in a
% Cartesian coordinate system) and gives the answer using prespecified
% basis vectors.

t = linspace(0,norm(P2-P1),1/step_size);
direction = (P2-P1)/norm(P2-P1);
points = [];
for j = t
    points = [points; (basis_vectors*(P1 + j*direction))'];
end

end
```

```
function parameters = PbTe
% Contains tight-binding parameters for PbTe
parameters = [-7.612
    -11.002
    3.195
    -0.237
    7.73
    7.73
    1.500
    0.428
    -0.474
    0.705
    0.633
    2.066
    -.0430
    -1.29
    0.835
    -1.59
    0.531
    -1.35
    0.668
    0.646e-9];
end
```

```
function parameters = SnTe
% Contains tight-binding parameters for SnTe.
parameters = [-6.578
    -12.067
    1.659
    -0.167
    8.38
    7.73
```

```
    0.592
    0.564
    -0.510
    0.949
    -0.198
    2.218
    -0.446
    -1.11
    0.624
    -1.67
    0.766
    -1.72
    0.618
    0.63e-9];
end
```

Finally we have the function `hamiltonian_matrix(kx,ky,kz,x)`, which generates the Hamiltonian matrix at reciprocal lattice point (k_x, k_y, k_z) for the material $Pb_{1-x}Sn_xTe$.

A.2 The tetragonal system

A.2.1 The bulk case

The main program is the following:

```
% Defines the tight-binding paramters that we use.
parameters = tetragonal_parameters;
t1A = parameters(1);
t1B = parameters(2);
t2A = parameters(3);
t2B = parameters(4);
t1prim = parameters(5);
t2prim = parameters(6);
tzprim = parameters(7);

% Sets the lattice constants to one.
a = 1;
c = 1;

% Defines the reciprocal lattice points that are of interest.
Gamma = [0,0,0];
M = [1/2,1/2,0];
A = [1/2,1/2,1/2];
Z = [0,0,1/2];

% Defined the reciprocal lattice vectors.
```

```
b1 = 2*pi/a*[1 0 0];
b2 = 2*pi/a*[0 1 0];
b3 = 2*pi/c*[0 0 1];

% Creates a matrix which tells us between which reciprocal latice points
% the band structure is to be calculated.
path_matrix = [Gamma; M; A; Z; Gamma];

% Creates a matrix which is used to get a plot with the correct scaling
% between the reciprocal lattice points.
distance_matrix = [];
for m = 1:size(path_matrix,1)-1
    distance_matrix = [distance_matrix norm(path_matrix(m,:)-...
        path_matrix(m+1,:))];
end
distance_matrix = distance_matrix/norm(distance_matrix,inf);

% This loop calculates and plots the band structure. In each loop the band
% structure between two of the chosen reciprocal lattice points is
% calculated.
s = 0;
xplace = [s];
for m = 1:size(path_matrix,1)-1
    % Evaluates at which points the Hamiltonian matrix should be
    % diagonalized.
    step_size = 0.001;
    P1 = path_matrix(m,:);
    P2 = path_matrix(m+1,:);
    [points,t] = line_between_points(P1',P2',[b1' b2' b3'],step_size);
    energies_Delta = [];

    % Calculates and diagonalizes the Hamiltonian matrix at appropriate
    % points in reciprocal space.
    for j = 1:size(points,1)
        k = points(j,:);
        kx = k(1);
        ky = k(2);
        kz = k(3);
          HA = 2*t1A*[cos(kx) 0; 0 cos(ky)] + 2*t2A*[cos(kx)*cos(ky)...
              sin(kx)*sin(ky); sin(kx)*sin(ky) cos(kx)*cos(ky)];
          HB = 2*t1B*[cos(kx) 0; 0 cos(ky)] + 2*t2B*[cos(kx)*cos(ky)...
              sin(kx)*sin(ky); sin(kx)*sin(ky) cos(kx)*cos(ky)];
          HAB = eye(2)*(t1prim+2*t2prim*(cos(kx)+cos(ky))+...
              tzprim*exp(1i*kz));
          H = [HA HAB; HAB' HB];
```

```
        E = eig(H);
        energies_Delta = [energies_Delta E];
    end

    % Plots the calculated band-structure.
    for j = 1:4
        figure(q)
        hold on
        plot(t/t(end)*distance_matrix(m)+s,energies_Delta(j,:),'-')
    end

    s = s+distance_matrix(m);
    xplace = [xplace, s];
    line([s,s],ylim,'Color','black')
end

% Makes the plot look nice.
xlim([0 s])
set(gca, 'xTick', xplace)
set(gca, 'xTickLabel', {'\Gamma','M','A','Z','\Gamma'})
```

The tight-binding parameters are defined in the following function:

```
function parameters = tetragonal_parameters
% Defines the tight-binding paratmeters for the tetragonal system.
parameters = [1
    -1
    1/2
    -1/2
    2.5
    0.5
    2];
end
```

A.2.2 The slab case

The main program is the following:

```
close all

% Defines the tight-binding parameters for the tetragonal system.
parameters = tetragonal_parameters;
t1A = parameters(1);
t1B = parameters(2);
t2A = parameters(3);
t2B = parameters(4);
```

```
t1prim = parameters(5);
t2prim = parameters(6);
tzprim = parameters(7);

a = 1;
c = 1;

% Defines the appropriate reciprocal lattice points.
Gamma_bar = [0 0 0];
M_bar = [1/2 1/2 0];
X_bar = [1/2 0 0];

% Creates a matrix which tells us between which points in the reciprocal
% space the band structure is to be calculated.
path_matrix = [Gamma_bar; M_bar; X_bar; Gamma_bar];

% Defines reciprocal lattice vectors.
b1 = 2*pi/a*[1 0 0];
b2 = 2*pi/a*[0 1 0];
b3 = 2*pi/c*[0 0 0];

q=1;
for layers=[2,5,10,15,20] % Defines the slab thickness in terms of the
    %number of AB-layers.

    % Creates a matrix which is used to get a plot with the correct scaling
    % between the reciprocal lattice points.
    distance_matrix = [];
    for m = 1:size(path_matrix,1)-1
        distance_matrix = [distance_matrix norm(path_matrix(m,:)-path_matrix(m+1,:))];
    end
    distance_matrix = distance_matrix/norm(distance_matrix);

    % This loop calculates and plots the band structure. In each loop the band
    % structure between two of the chosen reciprocal lattice points is
    % calculated.
    s = 0;
    xplace = [s];
    for m = 1:size(path_matrix,1)-1
        % Evaluates at which points the Hamiltonian matrix should be
        % diagonalized.
        step_size = 0.01;
        P1 = path_matrix(m,:);
        P2 = path_matrix(m+1,:);
        [points,t] = line_between_points(P1',P2',[b1' b2' b3'],step_size);
```

```
        % Finds the eigenvalues of the Hamiltonian matrix at the desired
        % k-points. This matrix is defined in the function.
        % tetragonal_hamiltonian.
        energies_Delta = [];
        for j = 1:size(points,1)
            k = points(j,:);
            H = tetragonal_hamiltonian(layers,k(1),k(2),k(3),a,c);
            E = eig(H);
            energies_Delta = [energies_Delta E];
        end

        % Plots the calculated bandstructure.
        for j = 1:size(energies_Delta,1)
            if layers~=20
                subplot(2,2,q)
            else
                figure(20)
            end
            hold on
            plot(t/t(end)*distance_matrix(m)+s,energies_Delta(j,:))
        end

        s = s+distance_matrix(m);
        xplace = [xplace, s];
        line([s,s],ylim,'Color','black')
    end

    % Makes the plots look nice.
    xlim([0,s])
    ylim([-8,8])
    set(gca, 'xTick', xplace)
    set(gca, 'xTickLabel', {'\Gamma','M','X','\Gamma'})
    title([num2str(layers),' layers'])
    ylabel('Energy')
    q=q+1
end
```

The Hamiltonian matrix is generated in the following function:

```
function Hslab = tetragonal_hamiltonian(layers,kx,ky,kz,a,c)
% Calulcates the Hamiltonian matrix for a slab of thickness layers at
% reciprocal lattice point (kx,ky,kz).

% Defines the tight-binding parameters.
parameters = tetragonal_parameters;
```

```
t1A = parameters(1);
t1B = parameters(2);
t2A = parameters(3);
t2B = parameters(4);
t1prim = parameters(5);
t2prim = parameters(6);
tzprim = parameters(7);

% Calculates the blocks that the slab matrix is built up from.
HA = 2*t1A*[cos(a*kx) 0; 0 cos(a*ky)] + 2*t2A*[cos(a*kx)*cos(a*ky)...
    sin(a*kx)*sin(a*ky); sin(a*kx)*sin(a*ky) cos(a*kx)*cos(a*ky)];
HB = 2*t1B*[cos(a*kx) 0; 0 cos(a*ky)] + 2*t2B*[cos(a*kx)*cos(a*ky)...
    sin(a*kx)*sin(a*ky); sin(a*kx)*sin(a*ky) cos(a*kx)*cos(a*ky)];
HAB = eye(2)*(t1prim+2*t2prim*(cos(a*kx)+cos(a*ky)));
HI = eye(2)*(tzprim*exp(1i*c*kz));
H0 = [HA HAB; HAB' HB];
Hinterlayer = [zeros(2) zeros(2); HI' zeros(2)];

% Puts H0 on the diagonal of the matrix Hslab.
A0 = repmat(H0,1,layers);
B0 = mat2cell(A0,size(H0,1),ones(1,layers)*size(H0,2));
Hslab = blkdiag(B0{:});

% Creates a new matrix Hpart12big that contains Hinterlayer on the
% subdiagonal.
A12 = repmat(Hinterlayer,1,layers-1);
B12 = mat2cell(A12,size(Hinterlayer,1),ones(1,layers-1)*...
    size(Hinterlayer,2));
Hpart12 = blkdiag(B12{:});
Hpart12big = [zeros(size(Hpart12,1),size(H0,2)) Hpart12;...
    zeros(size(H0,1),size(Hslab,2))];

% Hinterlayer is added to the subdiagonals of Hslab.
Hslab = Hslab + Hpart12big + Hpart12big';

end
```

B Some mathematical concepts

B.1 Anti-symmetric matrices

We will in many cases of this thesis be interested in anti-symmetric matrices, and thus we will discuss some important properties of those. Let A be a matrix which satisfies

$$A^T = -A, \tag{B.1}$$

where T denotes the transpose, is called an *anti-symmetric* matrix. Written in component form, if $A = (a_{ij})$, then $a_{ij} = -a_{ji}$ if A is anti-symmetric.

Now we make the following definition:

Definition B.1. [36] Let $A = (a_{ij})$ be an anti-symmetric $2n \times 2n$-matrix. The *Pfaffian* of A is given by

$$\mathrm{Pf}(A) = \sum_{\alpha} \mathrm{sgn}(\sigma_\alpha) a_{i_1 j_2} a_{i_2 j_2} \dots a_{i_n j_n}, \tag{B.2}$$

where the sum runs over all possible partitions α of the set $\{1, 2, \dots, 2n\}$ into non-intersecting pairs $\{i_k, j_k\}$, where $i_k < j_k$ and $k = 1, \dots, n$, and σ_α is the permutation

$$\begin{pmatrix} 1 & 2 & \dots & 2n-1 & 2n \\ i_1 & j_1 & \dots & i_n & j_n \end{pmatrix}. \tag{B.3}$$

One can show that the Pfaffian has the following properies

Proposition B.1. *[36] For an anti-symmetric $2n \times 2n$-matrix A and an arbitrary $2n \times 2n$-matrix B, we have*

$$(\mathrm{Pf}(A))^2 = \det(A), \tag{B.4}$$

and

$$\mathrm{Pf}(B^T A B) = \det(C)\mathrm{Pf}(A). \tag{B.5}$$

B.2 Principal bundles

A *fiber bundle* [37] is a collection (E, B, p, F), where E, B and F are topological spaces and p is a map $p : E \to B$ called the *projection map*. The space B is called the *base space*, E the *total space* and F the *fiber*.

We require that for every point $e \in E$ there is an open neighbourhood $U \subset B$ of $p(e)$ such that there exists a homeomorphism

$$\phi : p^{-1}(U) \to U \times F, \tag{B.6}$$

which satisfies

$$\mathrm{proj}_U \circ \phi = p_{p^{-1}(U)}, \tag{B.7}$$

where $\mathrm{proj}_U : U \times F$ is the natural projection onto U. The homomorphisms ϕ are called *local trivializations* of the fiber bundle.

Sometimes fiber bundles admit a map called the *global section*. Such a map $s : B \to E$ is continuous and has the property that $p \circ s$ is the identity. When it is not possible to define a global section, one can instead define *local sections*. This is simply a continuous map $s : U \to E$, where U is an open set in B and $p \circ s$ is the identity.

A *principal bundle* [38] is a special case of a fiber bundle with a continuous right action of a topological group G on the total space, $E \times G \to E$, such that the fibers of the fiber bundle are preserved and such that G acts freely and transitively on them. This means that each fiber of the fiber bundle is homeomorphic to the group G.

One concept, is the concept of a *connection*. This is something that gives us the notion of parallel transport on a bundle. More precisely, this is a way to connect fibers over nearby points. In the case of a principal bundle, the connection should be compatible with the action of the group G. We will however not go into the more technical bits of this description. An example of a connection is seen in equation (3.18).

References

[1] X.-G. Wen, "A theory of 2+ 1d bosonic topological orders," *National Science Review*, vol. 3, no. 1, pp. 68–106, 2016.

[2] L. D. Landau *et al.*, "On the theory of phase transitions," *Zh. eksp. teor. Fiz*, vol. 7, no. 19-32, 1937.

[3] D. C. Tsui, H. L. Stormer, and A. C. Gossard, "Two-dimensional magnetotransport in the extreme quantum limit," *Physical Review Letters*, vol. 48, no. 22, p. 1559, 1982.

[4] J. G. Bednorz and K. A. Müller, "Possible hightc superconductivity in the ba- la- cu- o system," *Zeitschrift für Physik B Condensed Matter*, vol. 64, no. 2, pp. 189–193, 1986.

[5] Y. Ando, "Topological insulator materials," *Journal of the Physical Society of Japan*, vol. 82, no. 10, p. 102001, 2013.

[6] M. H. Berntsen, "Consequences of a non-trivial band-structure topology in solids: Investigations of topological surface and interface states," Ph.D. dissertation, KTH Royal Institute of Technology, 2013.

[7] T. R. S. A. of Sciences. (2016) Scientific background on the nobel prize in physics 2016 - topological phase transitions and topological phases of matter. [Online]. Available: https://www.nobelprize.org/nobel_prizes/physics/laureates/2016/advanced-physicsprize2016.pdf

[8] L. Fu, "Topological crystalline insulators," *Physical Review Letters*, vol. 106, no. 10, p. 106802, 2011.

[9] J. C. Slater and G. F. Koster, "Simplified lcao method for the periodic potential problem," *Physical Review*, vol. 94, no. 6, p. 1498, 1954.

[10] J. J. Sakurai and J. J. Napolitano, *Modern quantum mechanics*. Pearson Higher Ed, 2014.

[11] G. David, *Introduction to Quantum Mechanics: Pearson New International Edition*. PE, 2005.

[12] H. Lüth, *Surfaces and interfaces of solid materials*. Springer Science & Business Media, 2013.

[13] T. Wehling, A. M. Black-Schaffer, and A. V. Balatsky, "Dirac materials," *Advances in Physics*, vol. 63, no. 1, pp. 1–76, 2014.

[14] M. P. Marder, *Condensed matter physics*. John Wiley & Sons, 2010.

[15] T. Ihn, *Semiconductor Nanostructures: Quantum states and electronic transport*. Oxford University Press, 2010.

[16] M. Nakahara, *Geometry, topology and physics*. CRC Press, 2003.

[17] L. Fu and C. L. Kane, "Time reversal polarization and a z 2 adiabatic spin pump," *Physical Review B*, vol. 74, no. 19, p. 195312, 2006.

[18] B. A. Bernevig and T. L. Hughes, *Topological insulators and topological superconductors*. Princeton University Press, 2013.

[19] R. Resta, "Electrical polarization and orbital magnetization: the modern theories," *Journal of Physics: Condensed Matter*, vol. 22, no. 12, p. 123201, 2010.

[20] J. E. Moore and L. Balents, "Topological invariants of time-reversal-invariant band structures," *Physical Review B*, vol. 75, no. 12, p. 121306, 2007.

[21] L. Fu and C. L. Kane, "Topological insulators with inversion symmetry," *Physical Review B*, vol. 76, no. 4, p. 045302, 2007.

[22] J. Thijssen, *Computational physics*. Cambridge university press, 2007.

[23] T. Heinzel, *Mesoscopic electronics in solid state nanostructures*. John Wiley & Sons, 2008.

[24] P.-O. Löwdin, "On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals," *The Journal of Chemical Physics*, vol. 18, no. 3, pp. 365–375, 1950.

[25] A. Szabo and N. S. Ostlund, *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation, 2012.

[26] L. Kantorovich, *Quantum theory of the solid state: an introduction*. Springer Science & Business Media, 2004, vol. 136.

[27] P. Lowdin, *Advances in Quantum Chemistry*, ser. Advances in Quantum Chemistry. Elsevier Science, 1988, no. v. 19.

[28] A. T. Paxton *et al.*, "An introduction to the tight binding approximation–implementation by diagonalisation," *NIC Series*, vol. 42, pp. 145–176, 2009.

[29] P. Dziawa, B. Kowalski, K. Dybko, R. Buczko, A. Szczerbakow, M. Szot, E. Łusakowska, T. Balasubramanian, B. M. Wojek, M. Berntsen *et al.*, "Topological crystalline insulator states in pb1- xsnxse," *Nature materials*, vol. 11, no. 12, pp. 1023–1027, 2012.

[30] M. Grundmann, *The Physics of Semiconductors: An Introduction Including Nanophysics and Applications*. Springer, 2015.

[31] C. S. Lent, M. A. Bowen, J. D. Dow, R. S. Allgaier, O. F. Sankey, and E. S. Ho, "Relativistic empirical tight-binding theory of the energy bands of gete, snte, pbte, pbse, pbs, and their alloys," *Superlattices and Microstructures*, vol. 2, no. 5, pp. 491–499, 1986.

[32] G. W. Stewart, *Matrix perturbation theory*. Citeseer, 1990.

[33] D. C. Lay, *Linear Algebra and Its Applications 4th Edition Low Cost Soft Cover IE Edition*. Pearson Education Inc., 2011.

[34] S. Noschese, L. Pasquini, and L. Reichel, "Tridiagonal toeplitz matrices: properties and novel applications," *Numerical linear algebra with applications*, vol. 20, no. 2, pp. 302–326, 2013.

[35] A. Greenbaum, *Iterative methods for solving linear systems*. SIAM, 1997.

[36] E. of Matematics. Pfaffian. [Online]. Available: http://www.encyclopediaofmath.org/index.php?title=Pfaffian&oldid=35223

[37] T. Rowland. Fiber Bundle. From MathWorld–A Wolfram Web Resource, created by Eric W. Weisstein. [Online]. Available: http://mathworld.wolfram.com/FiberBundle.html

[38] T. Rowland. Principal Bundle. From MathWorld–A Wolfram Web Resource, created by Eric W. Weisstein. [Online]. Available: http://mathworld.wolfram.com/PrincipalBundle.html