DET NATURVIDENSKABELIGE FAKULTET
KØBENHAVNS UNIVERSITET

Understanding the low energy physics of bismuth selenide: A three-dimensional topological insulator

Author:
Mathias Rosdahl Jensen
Supervisors:
Jens Paaske
Anders Mathias Lunde

KØBENHAVNS UNIVERSITET

Master's Thesis • The Niels Bohr Institute • 1 September 2013 – 25 July 2014

Abstract

In this thesis, we give a thourough investigation of the basic physics of bismuth selenide, a recently discovered three-dimensional topological insulator. We give a detailed and pedagogical introduction to group theory, describing the symmetry operations of the crystal lattice, in order to construct the minimal effective model, describing the topological features of bismuth selenide. Qualitatively, we discuss the physical principles of the band structure around the Fermi level, which is found to consist of linear combinations of p-orbitals. Specifically, we see that a strong spin-orbit coupling leads to a band inversion. This band inversion gives rise to a non-trivial topology. Within this model, we calculate the topological surface states by imposing hard-wall boundary conditions. For a single isolated surface we find the conditions on the parameters of the model, for the existence of surface states. We analytically find the spectrum and wave functions of the surface states. These have a Dirac-like spectrum, and a helical spin structure. In a thin film, the overlap of wavefunctions on opposite surfaces, leads to a gap in the spectrum. We discuss the dependence of the gap on the thickness, as well as the parameters of the model and compare to experimental measurements of the gap. For a thin film, the spin structure is dependent on position. The helical spin structure, gets opposite vorticity on the two surfaces, which is a result of the inversion symmetry of the crystal.

Contents

Introduction

1.1 Historical background for the dicovery of 3D topological insulators

In 1980 von Klitzing *et al.* discovered the integer quantum Hall effect [1], where a large magnetic field, gives rise to a quantized Hall conductance, σ_{xy}. The Hall conductance is the ratio, between the current through the material and the voltage across the perpendicular direction. He was awarded the Nobel Prize in 1985 for this discovery. In 1982 Thouless, Kohmoto, Nightingale and den Nijs found that this phenomenon was related to topology. Specifically the topology of the Hilbert space of wave functions in a quantum Hall system. They defined an integer *topological invariant*, which characterizes the topology of the Hilbert space. They showed that the Hall conductivity became equal to this integer times $\frac{e}{h}$. This was the first example of a *topological quantum number*. But for a long time no one cared about topology in physics.

Many years later the two dimensional carbon material graphene became famous after it was discovered in 2004 [2]. The material attracted a huge amount of attention, and in 2010 Andre Geim and Konstantin Novoselov recieved the Nobel Prize *"for groundbreaking experiments regarding the two-dimensional material graphene"*. It was proposed that graphene should be a 2D topological insulator [3], but it required an urealistically strong spin-orbit coupling [4]. But the concept of topology, was back in the game of physics. In 2006 Bernevig, Hughes and Zhang proposed that a mercury telluride (HgTe) quantum well should be a topological insulator [5]. In 2007 this proposal was experimentally verified by König *et al.* [6].

Now with the discovery of the first two-dimensional topological insulator, naturally, a question arises; Does something equivalent exist in three dimensions? This is indeed the case and the subject of the present thesis.

1.2 What is a 3D topological insulator?

As the name suggests, a topological insulator, is related to an insulator. In fact, a topological insulator is insulating inside the material, because of a gap in the spectrum. But on the surface, exotic *topological surface states* exist. Energetically these states lie within the bulk band gap. Therefore, the surface of a topological insulator is conducting. One interesting feature of the surface states is that they behave as massless Dirac fermions, just like in graphene. But in contrast to graphene, the surface states in 3D topological insulators exhibit a *spin momentum locking*, where the spin is always perpendicular to the momentum. This spin structure, shown in figure 1.1 is sometimes called *helical*. The helical spin structure can rotate in clock-wise, or counter clock-wise fashion, which are refered to as negative and positive vorticity, respectively.

The simplest model, that gives rise to these *helical* states, is the Dirac-like Hamiltonian:

$$H_{\text{Surface}}(\mathbf{k}) = v_F(k_y\sigma_x - k_x\sigma_y) \tag{1.1}$$

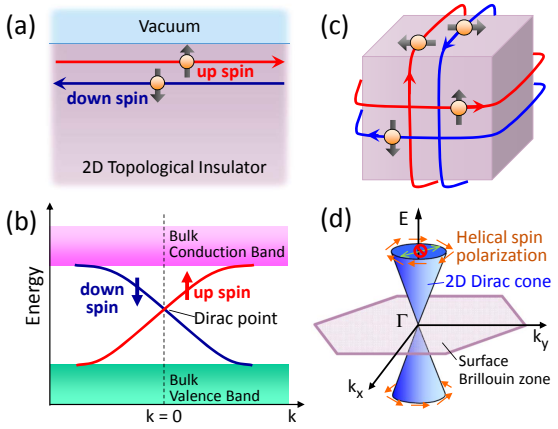


Figure 1.1: (a) Real space picture of the helical edge state of a 2D topological insulator. (b) Energy dispersion for a 2D topological insulator. (c) Real space picture of the helical surface states of a 3D topological insulator. (d) Energy dispersion and spin structure of the surface states on a 3D topological insulator. Adopted from [4].

This model is widely used to describe the surface states of a 3D topological insulator. Note that this is a 2D model describing only the surface states, completely ignoring the three dimensional nature of these states. The eigenstates of this Hamiltonian are:

$$\psi = \frac{1}{\sqrt{2}} \begin{pmatrix} -i \\ \pm \frac{k_+}{|\mathbf{k}|} \end{pmatrix} \quad E = \pm v_F |\mathbf{k}| \tag{1.2}$$

Giving a linear dispersion, just like massless relativistic fermions. We also get the helical spin structure by calculating the expectation value of the spin:

$$\langle \sigma_x \rangle_{E=\pm v_F|\mathbf{k}|} = \mp \frac{k_y}{2|\mathbf{k}|} \quad \langle \sigma_y \rangle_{E=\pm v_F|\mathbf{k}|} = \pm \frac{k_x}{2|\mathbf{k}|} \tag{1.3}$$

1.3 Bismuth Selenide: Well known, but brand new

Bismuth selenide (Bi_2Se_3) is a material, which is widely known and used because of its thermoelectric properties. But recently the development of the field of topolological physics has caused a renewed interest in the material. In 2009 Bi_2Se_3 and the related materials Bi_2Te_3 and Sb_2Te_3 were predicted to be three dimensional topological insulators.

Experiments using angle resolved photo emission spectroscopy (ARPES), can map out the spectrum of the surface states, which is one of the most important ways to verify the predicted existence of surface states. In 2009 the topological nature of Bi_2Se_3 was verified by this technique [7].

In 3D topological insulators transport measurements have not been so succesfull yet, since the bulk often turns out to be conducting as well, due to impurities. One way of reducing this problem is simply to make a thin film, such that the bulk is less important. Therefore, thin films are important to understand, especially how the surface states on opposite surfaces, affect each other.

1.4 Outline of the thesis

In this thesis, Bi_2Se_3 is studied in detail. We begin with a chapter on group theory, discussing the neccessary concepts and methods, which are used to construct the simplest model for Bi_2Se_3. This model is constructed in chapter 3, after a detailed description of the basic physical principles of the electronic structure of Bi_2Se_3. Thereafter, we dive into the study of surface states within this model. First, we will investigate a single isolated surface in chapter 4, where much analytical progress can be made. This gives an understanding, which serves as a great starting point for investigating a thin film in chapter 5. Finally the results are summarized in chapter 6

Group theory

In this chapter, we will introduce the necessary group theoretical concepts needed to derive the effective model for Bi_2Se_3. We will focus more on how to use the methods in physical systems than on rigorous proofs, though some important theorems will be proved to give an insight in the mathematical framework. In physical applications, the groups will be groups consisting of symmetry operations. Throughout this chapter we will use the group of symmetries of an equilateral triangle as an example, since it has physical application on its own and is part of the symmetry group of Bi_2Se_3, and hence it will be useful to have this in mind when introducing the various concepts.

2.1 Abstract groups

A group is defined as a set of elements $\mathcal{G} = \{A, B, C, ...\}$ for which some kind of multiplication is defined. This multiplication has to satisfy the following rules:

1. The group is closed under multiplication, i.e. $AB \in \mathcal{G}$ for any $A, B \in \mathcal{G}$.
2. The associative law holds, i.e $A(BC) = (AB)C$, for any elements $A, B, C \in \mathcal{G}$. This means that that the notation ABC is unambiguous.
3. There exist a unit element $E \in \mathcal{G}$ such that $EA = AE = A$ for any $A \in \mathcal{G}$.
4. To each element $A \in \mathcal{G}$ there exist an inverse element A^{-1} such that $AA^{-1} = A^{-1}A = E$.

A very simple (and uninteresting) example of a group is just $\mathcal{G} = \{1\}$, containing only one element. Another often used group is the set of positive real numbers, containing an infinite number of elements, but is often not referred to as group since it has a lot more structure. The number of elements in a group is called the *order* of the group. In the following we will focus on groups of finite order.

In physics the typical groups considered is a set of symmetry transformations, that leaves a physical system invariant. In this case the group multiplication AB is simply the transformation obtained by first performing the transformation B and then A. It should be obvious now that the set of symmetry operations on a system is group, but let us just check the four points in the definition;

1. If a system is invariant under the transformations A and B it should also be invariant under AB, i.e. doing one transformation and then the other.
2. Associativity comes directly by the definition of multiplication.
3. The identity transformation will always leave a system invariant.
4. If a transformation leaves a system invariant, then doing the inverse should also leave it invariant.

Unlike the group of positive real numbers, a general group is not necessarily commutative, i.e. in general $AB \neq BA$, as is the case for a group containing rotations around different axes.

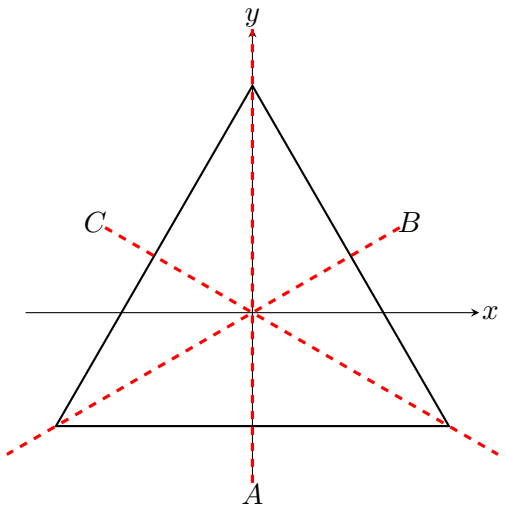


Figure 2.1: The group of symmetry operations on the equilateral triangle consists of the identity, two rotations around the z axis by 120 degrees (clockwise and counter-clockwise) and three rotations of 180 degrees around the axes A, B and C, indicated by the red dashed lines.

As an example let us consider the group called D_3 that leaves an equilateral triangle invariant, see figure 2.1. This group consists of the elements:

- The identity element which is usually denoted by E.
- Two rotations by 120 degrees around an axis perpendicular to the triangle, denoted by D (clockwise) and F (counter-clockwise).
- Three rotations of 180 degrees around the medians of the triangle, denoted by A, B and C.

By performing successive transformations of the group, we can verify that the group is indeed closed. The results of performing two successive transformations can be summarized in a multiplication table, see table 2.1[1]. The product AB is then read off as the element at the row labeled by A and the column labeled by B, and we see $AB = D$. We also immediately see that $BA = F \neq AB$, hence this group is not commutative.

Observe that each row or column contains each element of the group exactly once. This is no accident, it is a general theorem called the rearrangement theorem.

Rearrangement theorem: For any finite group $\mathcal{G} = \{A_1, ..., A_n\}$ the sequence:

$$A_1 A_k, ..., A_n A_k \tag{2.1}$$

contains each group element A_i exactly once, for any group element A_k.

Proof. For any A_k and A_i in the group there exists $A_r = A_i A_k^{-1} \in \mathcal{G}$ since all elements has an inverse and $\mathcal{G}$ is closed under multiplication. Hence $A_i = A_r A_k$ and thus appears at least

[1] If you try to reproduce this (or something similar) it is highly recommended to use an object which is *not* symmetrical (or label the corners of the triangle), else all transformations just leaves the object invariant.

	E	A	B	C	D	F
E	E	A	B	C	D	F
A	A	E	D	F	B	C
B	B	F	E	D	C	A
C	C	D	F	E	A	B
D	D	C	A	B	F	E
F	F	B	C	A	E	D

Table 2.1: Multiplication table of the symmetry group of an equilateral triangle. In agreement with the rearrangement theorem each row (and column) contain each element exactly once.

once in the above sequence. But since the sequence contains the same number of elements as the group, each element must appear exactly once. □

2.2 Classes

We will now see how the elements of a group can be divided into *classes* of elements, using the concept of *conjugation* of elements.

Two elements $A, B \in \mathcal{G}$ is said to be *conjugate* to each other if

$$A = XBX^{-1} \tag{2.2}$$

for some element $X \in \mathcal{G}$. If A and B are both conjugate to C, then they are also conjugate to each other:

$$\begin{aligned} C = XAX^{-1} = YBY^{-1} \\ \Leftrightarrow A = X^{-1}YBY^{-1}X \Leftrightarrow A = X^{-1}YB(X^{-1}Y)^{-1} \end{aligned} \tag{2.3}$$

Here we used that the inverse of a product is the product of the inverse in the reverse order $(X^{-1}Y)^{-1} = Y^{-1}X$, which can easily be shown.

This means that we can divide the elements of a group into *classes*, which are sets of mutually conjugate elements, i.e. in a group $\mathcal{G} = \{A_1, ..., A_n\}$ the class containing the element A_k is the elements:

$$A_1 A_k A_1^{-1}, ..., A_n A_k A_n^{-1}. \tag{2.4}$$

Note that some of these elements might be equal. The identity element is for example always its own class since $A_i E A_i^{-1} = E$. In a commutative group each element is its own class, since $A_i A_k A_i^{-1} = A_i A_i^{-1} A_k = A_k$.

2.2.1 Classes of D_3

In the group of symmetries of the equilateral triangle, there are three classes $\mathcal{C}_1 = \{E\}$, $\mathcal{C}_2 = \{A, B, C\}$ and $\mathcal{C}_3 = \{D, F\}$. This can be shown simply by looking at the multiplication table, and computing all the conjugates, but it can also be realized from physical reasoning. A, B, C are all rotations of 180 degrees around different axes, and D and F transforms these

axes into each other, so for example rotation around B can be obtained by first rotating the B-axis into the A-axis using F then rotating around the A-axis, and then finally rotating back $F^{-1} = D$. And as seen on the multiplication table $B = DAF$. But in this way of thinking a class is not some abstract mathematical concept, we see that we actually divides the group into sets of elements which are physically similar. This kind of reasoning is very useful when determining the classes of a physical symmetry group, since it can be very tedious to work out all the products in the multiplication table.

2.3 Representations

By a *representation*, Γ of an abstract group $\mathcal{G} = \{E, A_1, ..., A_n\}$, we mean a set of square matrices, with one matrix *representing* each element A_i of the group, denoted by $\Gamma(A_i)$. Note that these do not have to be different matrices for each element, the only requirement is that they fulfill the same multiplication table, i.e.

$$\Gamma(A)\Gamma(B) = \Gamma(AB) \quad \forall \quad A, B \in \mathcal{G} \tag{2.5}$$

The representation is of course a group itself. If all the representation matrices are different then the order of this group is the same as the original one, and the representation is said to be *true* or *faithful*. All groups have the so-called identical representation where all the elements are simply represented by the number 1, which is of course *unfaithful* unless the group is only the identity element.

If we have a representation $\Gamma(A)$ we can form an infinity of new representations by similarity transformations $\Gamma'(A) = M\Gamma(A)M^{-1}$

$$\begin{aligned} \Gamma'(A)\Gamma'(B) &= M\Gamma(A)M^{-1}M\Gamma(B)M^{-1} = M\Gamma(A)\Gamma(B)M^{-1} \\ &= M\Gamma(AB)M^{-1} = \Gamma'(AB) \end{aligned} \tag{2.6}$$

These representations, related by similarity transformations, are said to be *equivalent*, since they are basically just stated in a different basis. The set of all equivalent representations is sometimes called an equivalence class of representations.

Another way to get a new representation from two (or more) representations is to form the matrices:

$$\Gamma(A) = \begin{pmatrix} \Gamma^{(1)}(A) & 0 \\ 0 & \Gamma^{(2)}(A) \end{pmatrix} \tag{2.7}$$

which clearly is a representation. This is called a *reducible* representation, since it can be reduced to the representations $\Gamma^{(1)}$ and $\Gamma^{(2)}$. But this representation is also equivalent to one where the rows and columns have been mixed around, and is no longer on block form. A representation is reducible if it is equivalent to a representation where all the matrices are on block form. Usually the structure of a reducible representation is given by writing it as a sum of the representations it consists of in block form. In the above example one would write $\Gamma = \Gamma^{(1)} + \Gamma^{(2)}$. Note that this notation is not usual addition of matrices. An irreducible

representation is then a representation, which cannot be brought on block form by a similarity transformation.

The great orthogonality theorem. Consider all inequivalent, irreducible unitary[2] representations $\Gamma^{(i)}$ of a group $\mathcal{G}$ then:

$$\sum_{A\in\mathcal{G}} \Gamma^{(i)}(A)^*_{\mu\nu}\Gamma^{(j)}(A)_{\alpha\beta} = \frac{h}{l_i}\delta_{ij}\delta_{\mu\alpha}\delta_{\nu\beta}, \tag{2.8}$$

where the sum is over all elements in the group, $\Gamma^{(i)}(A)_{\mu\nu}$ is the $\mu\nu$ matrix element of the i'th representation of the element A. h is the order of the group and l_i is the dimension of the i'th representation. For a proof see [8]. This theorem can be interpreted geometrically, as orthogonality between a set of vectors in a vector space of dimension h, i.e. we can consider each matrix element of each representation as a vector $\Gamma^{(i)}_{\mu\nu} = (\Gamma^{(i)}(A_1)_{\mu\nu}, \Gamma^{(i)}(A_2)_{\mu\nu}, ..., \Gamma^{(i)}(A_h)_{\mu\nu})$ with one component for each element in the group. The vectors are labeled by three indices, the representation index and the two matrix indices, so the set consists of $\sum_i l_i^2$ vectors. Since the maximal possible number of orthogonal vectors is equal to the dimension of the vector space we have $\sum_i l_i^2 \leq h$. Furthermore it can be shown that equality always holds ([8] p. 31):

$$\sum_i l_i^2 = h \tag{2.9}$$

2.3.1 Representations of D_3

Here we will construct explicitly the irreducible representations of the group of the equilateral triangle. Of course we have the one dimensional identical representation, where all elements are represented by the number 1.

The elements of the group act linearly on the vector (x, y) and thus, if we simply construct the matrices transforming this vector, we get a 2 dimensional representation of D_3. These matrices are found to be:

$$\begin{aligned}
\Gamma^{(3)}(E) &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} & \Gamma^{(3)}(A) &= \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} & \Gamma^{(3)}(B) &= \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix} \\
\Gamma^{(3)}(C) &= \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & \frac{1}{2} \end{pmatrix} & \Gamma^{(3)}(D) &= \begin{pmatrix} -\frac{1}{2} & \frac{\sqrt{3}}{2} \\ -\frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix} & \Gamma^{(3)}(F) &= \begin{pmatrix} -\frac{1}{2} & -\frac{\sqrt{3}}{2} \\ \frac{\sqrt{3}}{2} & -\frac{1}{2} \end{pmatrix}
\end{aligned} \tag{2.10}$$

Now one might wonder whether this representation is reducible. If it is reducible then all of the matrices could be diagonal by some similarity transformation, i.e. there exists some matrix M such that $M\Gamma^{(3)}(X)M^{-1}$ is diagonal for all X in D_3. This corresponds simply to a change of coordinate system in the plane, but a rotation by $\frac{2\pi}{3}$ will always mix the coordinates and the matrices can therefore not be made diagonal. Thus this is an irreducible representation.

If we also consider a z-axis perpendicular to the triangle, then the z coordinate will be invariant under E, D, F but will change sign under the rotations A, B, C, hence we can construct a new representation governing the transformation properties of the z-coordinate, given

[2] A unitary representation is simply a representation of unitary matrices. And any representation with nonvanishing determinant, is equivalent to a unitary representation, [8]

by:

$$\begin{aligned}\Gamma^{(2)}(E) = \Gamma^{(2)}(D) = \Gamma^{(2)}(F) = 1 \\ \Gamma^{(2)}(A) = \Gamma^{(2)}(B) = \Gamma^{(2)}(C) = -1\end{aligned} \tag{2.11}$$

Now we have three inequivalent, irreducible representations, since a one dimensional representation is always irreducible. And since we have the equation:

$$\sum_i l_i^2 = 6, \tag{2.12}$$

then these three are all the inequivalent, irreducible representations of D_3.

2.4 Characters

Since all representations are equivalent to an infinity of representations by similarity transformations, which may look very different, it would be useful to be able to characterize an equivalence class of representations rather than just a single representation. This will enable us to quickly check whether a representation is reducible and if two representations are equivalent. This can be done using the *character* of a representation, which is a set of h numbers $\chi^{(j)}(A_1), ..., \chi^{(j)}(A_n)$, where j is the index of the representation. This is given by:

$$\chi^{(j)}(A_i) = \operatorname{Tr} \Gamma^{(j)}(A_i) \tag{2.13}$$

Since the trace of a matrix is invariant under similarity transformations, this character is indeed the same for equivalent representations. This property also shows that all elements in one class have the same characters and it makes sense to simply talk about the characters of a class $\chi^{(j)}(\mathcal{C})$.

The great orthogonality theorem for characters. The characters of the irreducible representations of a group $\mathcal{G}$ obey the orthogonality relation:

$$\sum_{A\in\mathcal{G}} \chi^{(i)}(A)^* \chi^{(j)}(A) = \sum_k N_k \chi^{(j)}(\mathcal{C}_k)^* \chi^{(k)}(\mathcal{C}_k) = h\delta_{ij} \tag{2.14}$$

the first equality is just using that the character is the same for elements in same class. N_k is the number of elements in th class $\mathcal{C}_k$.

Proof. We start by the great orthogonality theorem:

$$\sum_{A\in\mathcal{G}} \Gamma^{(i)}(A)^*_{\mu\nu} \Gamma^{(j)}(A)_{\alpha\beta} = \frac{h}{l_i} \delta_{ij}\delta_{\mu\alpha}\delta_{\nu\beta}, \tag{2.15}$$

Then we set $\mu = \nu$ and $\alpha = \beta$ and sum over both ν and β:

$$\begin{aligned}\sum_{\nu,\beta}\sum_{A\in\mathcal{G}} \Gamma^{(i)}(A)^*_{\nu\nu} \Gamma^{(j)}(A)_{\beta\beta} &= \sum_{\nu,\beta} \frac{h}{l_i}\delta_{ij}\delta_{\nu\beta}\delta_{\nu\beta} \\ \Leftrightarrow \sum_{A\in\mathcal{G}} \chi^{(i)}(A)^* \chi^{(j)}(A) &= \frac{h}{l_i}\delta_{ij}\sum_\nu \delta_{\nu\nu} = h\delta_{ij}\end{aligned} \tag{2.16}$$

□

This orthogonality relation is extremely important, since it shows the power of characters. With this at hand it becomes easy to check whether a representation is reducible or not. If it is reducible, one can find the representations it is composed of. The character of a reducible representation Γ is simply the sum of the characters it is composed of, which is easy to see when it is on block form. Thus we can have:

$$\chi(A) = \sum_j a_j \chi^{(j)}(A) \tag{2.17}$$

where the sum is over irreducible representations, and a_j is an integer denoting the number of times the j'th irreducible representation occurs in Γ. Now we can use the orthogonality theorem for characters to find the integers a_j:

$$\begin{aligned} \sum_{A\in\mathcal{G}} \chi(A)\chi^{(i)}(A)^* = \sum_{A\in\mathcal{G}} \sum_j a_j \chi^{(j)}(A)\chi^{(i)}(A)^* = \sum_j a_j h\delta_{ij} = h a_i \\ \Leftrightarrow a_i = h^{-1} \sum_{A\in\mathcal{G}} \chi(A)\chi^{(i)}(A)^* = \sum_k \chi(\mathcal{C}_k)\chi^{(i)}(\mathcal{C}_k)^* \end{aligned} \tag{2.18}$$

This shows that the integers a_j are uniquely determined and can easily be found from the characters of the irreducible representations.

The great orthogonality theorem for characters shows that the characters can be seen as a set of orthogonal vectors, one for each irreducible representation. The dimension of this vector space is equal to the number of classes, and hence the number of irreducible representations cannot exceed the number of classes. In fact, it can be shown that equality always holds:

$$\text{Number of classes} = \text{Number of inequivalent irreducible representations} \tag{2.19}$$

This can be very useful since the number of classes usually can be determined by physical reasoning, while it can be a tedious exercise to figure out all the representations of a group.

As we have seen now a lot of information about a group is given by the characters of the irreducible representations. A convenient way to display these characters is to construct the *character table*. The columns are labeled by the classes usually with the number of elements in front and the rows are labeled by the irreducible representations. See table 2.3 for the case of the symmetry group of an equilateral triangle. The first row is just ones, since the first representation is the identical representation. The first column is just given by the dimension of the representation since the first class is just the identity element.

2.4.1 Construction of character tables

In most cases the character table of a group can be constructed by following these steps:

1. The number of irreducible representations is equal to the number of classes which can quite easily be determined by physical reasoning.

2. The dimensionalities l_i of the representations can in most cases be uniquely determined

by $\sum_i l_i^2 = h$. This will determine the first column of the character table since the $\chi^{(i)} = l_i$ (the trace of an identity matrix is simply equal to its dimension). The first row is also given already, since any group has the identity representation, where all elements are represented by the number 1.

3. The great orthogonality theorem for characters; rows must be orthogonal and normalized to h, using the weighting factor N_k, the number of elements in the class $\mathcal{C}_k$, i.e.:

$$\sum_k \chi^{(i)}(\mathcal{C}_k)^* \chi^{(j)}(\mathcal{C}_k) N_k = h\delta_{ij} \tag{2.20}$$

4. The columns must be orthogonal and normalized to $\frac{h}{N_k}$, i.e.:

$$\sum_i \chi^{(i)}(\mathcal{C}_k)^* \chi^{(i)}(\mathcal{C}_l) = \frac{h}{N_k}\delta_{kl} \tag{2.21}$$

This follows from the great orthogonality theorem for characters.

2.4.2 Character table of D_3

Now we will see how these rules can be used to determine the character table, without even writing down the multiplication table. Step 1 and 2 immediately gives all but the lower right 4 characters, see table 2.2. Now we will determine a, b, c and d by the steps 3 and 4. Step 3

	$\{E\}$	$2\mathcal{C}_3$	$3\mathcal{C}_2$
$\Gamma^{(1)}$	1	1	1
$\Gamma^{(2)}$	1	b	a
$\Gamma^{(3)}$	2	d	c

Table 2.2

gives the equations:

$$1 + 3|a|^2 + 2|b|^2 = 6 \tag{2.22}$$
$$4 + 3|c|^2 + 2|d|^2 = 6 \tag{2.23}$$
$$1 + 3a + 2b = 0 \tag{2.24}$$
$$2 + 3c + 2d = 0 \tag{2.25}$$
$$2 + 3a^*c + 2b^*d = 0 \tag{2.26}$$

while step 4 gives the equations:

$$1 + |a|^2 + |c|^2 = 2 \tag{2.27}$$
$$1 + |b|^2 + |d|^2 = 3 \tag{2.28}$$
$$1 + a + 2c = 0 \tag{2.29}$$
$$1 + b + 2d = 0 \tag{2.30}$$
$$1 + a^*b + c^*d = 0 \tag{2.31}$$

Now we can also use that since $\Gamma^{(2)}$ is a one dimensional representation so the representation matrices are simply the characters. Since $\mathcal{C}_2$ are rotations by π they are their own inverse, so $a = \chi^{(2)}(\mathcal{C}_2) = \pm 1$. Put this into equation (2.27) and we get $c = 0$. Putting this into equation (2.29) gives $a = -1$, equation (2.31) now gives $b = 1$ and at last equation (2.30) gives $d = -1$. Now we have the full character table, see 2.3.

	$\{E\}$	$2\mathcal{C}_3$	$3\mathcal{C}_2$
$\Gamma^{(1)}$	1	1	1
$\Gamma^{(2)}$	1	1	-1
$\Gamma^{(3)}$	2	-1	0

Table 2.3

2.5 Direct product groups

As we saw above, even though a character table can be determined from these rules, it takes some effort. For groups of higher order it becomes more tedious, maybe not even possible. But if a group can be divided into two types such that each element of one commutes with all elements of the other, then it can be written as a *direct product group*. This could for example be the group of rotations and the group of the inversion operation.

For two groups $\mathcal{G}_A = \{E, A_2, ..., A_{h_A}\}$ and $\mathcal{G}_B = \{E, B_2, ..., B_{h_B}\}$, where any element of $\mathcal{G}_A$ commutes with all elements of $\mathcal{G}_B$ we define the direct product of $\mathcal{G}_A$ and $\mathcal{G}_B$ as:

$$\mathcal{G}_A \times \mathcal{G}_B = \{E, A_2, ..., A_{h_A}, EB_2, A_2B_2, ..., A_{h_A}B_2, ..., EB_{h_B}, A_2B_{h_B}, ..., A_{h_A}B_{h_B}\} \tag{2.32}$$

This is a group of order $h_A h_B$. It is closed under multiplication since elements of different groups commutes.

Now we will try to find the irreducible representations of the direct product group. An educated guess would be to use the direct product (or kronecker product) of the matrices of the representations of the original groups. For matrices M and N of dimensions a, b and c, d the direct product $\mathbf{M} \otimes \mathbf{N}$ is a ac by bd matrix, in block form given by:

$$\mathbf{M} \otimes \mathbf{N} = \begin{pmatrix} M_{11}\mathbf{N} & \cdots & M_{1b}\mathbf{N} \\ \vdots & \ddots & \vdots \\ M_{a1}\mathbf{N} & \cdots & M_{ab}\mathbf{N} \end{pmatrix} \tag{2.33}$$

For 4 matrices $\mathbf{M}, \mathbf{N}, \mathbf{L}, \mathbf{K}$ of dimensions such that the matrix products $\mathbf{ML}$ and $\mathbf{NK}$ exists, the so-called *mixed product property* holds:

$$(\mathbf{M} \otimes \mathbf{N})(\mathbf{L} \otimes \mathbf{K}) = (\mathbf{ML}) \otimes (\mathbf{NK}) \tag{2.34}$$

With this property we can show that the direct product matrices of a representation of $\mathcal{G}_A$ with a representation of $\mathcal{G}_B$, $\Gamma^{\mathcal{G}_A \times \mathcal{G}_B}(A_i B_j) = \Gamma^{\mathcal{G}_A}(A_i) \otimes \Gamma^{\mathcal{G}_B}(B_j)$ forms a representation of

the direct product group:

$$\begin{aligned}\Gamma^{\mathcal{G}_A\times\mathcal{G}_B}(A_iB_j)\Gamma^{\mathcal{G}_A\times\mathcal{G}_B}(A_kB_l) &= \left[\Gamma^{\mathcal{G}_A}(A_i)\otimes\Gamma^{\mathcal{G}_B}(B_j)\right]\left[\Gamma^{\mathcal{G}_A}(A_k)\otimes\Gamma^{\mathcal{G}_B}(B_l)\right] \\ = \left[\Gamma^{\mathcal{G}_A}(A_i)\Gamma^{\mathcal{G}_A}(A_k)\right]\otimes\left[\Gamma^{\mathcal{G}_B}(B_j)\Gamma^{\mathcal{G}_B}(B_l)\right] &= \Gamma^{\mathcal{G}_A}(A_iA_k)\otimes\Gamma^{\mathcal{G}_B}(B_jB_l) \\ &= \Gamma^{\mathcal{G}_A\times\mathcal{G}_B}(A_iA_kB_jB_l) \end{aligned} \tag{2.35}$$

In [8] it is shown that the irreducible representations of a direct product group $\mathcal{G}_A \times \mathcal{G}_B$ is precisely the direct products of the irreducible representations of $\mathcal{G}_A$ and $\mathcal{G}_B$. The classes of the direct product groups is easily found from the classes of the original groups, each pair of classes (one from each group) form a class in the product group. This can be seen by using that the elements of the two groups commute. For a general element in $\mathcal{G}_A \times \mathcal{G}_B$ A_jB_k we have:

$$XA_jB_kX^{-1} = XA_jX^{-1}B_k \text{ if } X \in \mathcal{G}_A \tag{2.36}$$

$$XA_jB_kX^{-1} = A_jXB_kX^{-1} \text{ if } X \in \mathcal{G}_B \tag{2.37}$$

So the total number of classes is the product of the numbers of classes in the two original groups, in agreement with the number of irreducible representations. The character of a direct product representation $\chi^{A\times B}(A_iB_j)$ is simply the product of the characters of the original representation $\chi^A(A_i)\chi^B(B_j)$, which just follows from the property of the kronecker product $\operatorname{Tr} A \otimes B = \operatorname{Tr} A \operatorname{Tr} B$. This allows us to construct the character table of $\mathcal{G}_A \times \mathcal{G}_B$ if the character tables for $\mathcal{G}_A$ and $\mathcal{G}_B$ are known.

2.5.1 Character table of D_{3d}

The elements of D_3 all commutes with the inversion operator I, which transforms $\mathbf{x} \to -\mathbf{x}$. Therefore we can construct the direct product group $D_{3d} = D_3 \times I$, where I denotes the group containing only I and the identity. The character table of the group of the inversion operator is easily constructed; first row and column are ones due to the two first rules, and the last element must then be one because of the orthogonality relation. Using the character tables of

	$\{E\}$	I
$\Gamma^{(+)}$	1	1
$\Gamma^{(-)}$	1	-1

Table 2.4: Character table for the group of the inversion operator.

D_3 and I we see that the character table of D_{3d} is simply four copies of the charactertable of D_3, but with a minus sign on one of them, see table 2.5.

2.6 Basis functions

Now it is also interesting to let symmetry operations operate on functions. This is useful since eventually we will use group theory to characterize quantum mechanical wave functions. Consider a group of coordinate transformations $\mathcal{G} = \{E, A_1, ...\}$ which transforms the coordinates

	$\{E\}$	$2\mathcal{C}_3$	$3\mathcal{C}_2$	I	$2\mathcal{C}_3 I$	$3\mathcal{C}_2 I$
$\Gamma^{(1+)}$	1	1	1	1	1	1
$\Gamma^{(2+)}$	1	1	-1	1	1	-1
$\Gamma^{(3+)}$	2	-1	0	2	-1	0
$\Gamma^{(1-)}$	1	1	1	-1	-1	-1
$\Gamma^{(2-)}$	1	1	-1	-1	-1	1
$\Gamma^{(3-)}$	2	-1	0	-2	1	0

Table 2.5: Character table for the group D_{3d}

$\mathbf{x} \to \mathbf{x}' = A_i\mathbf{x}$. Here A_i is not necessarily a matrix it is just some transformation of the coordinate, which could also include a translation. Now we introduce a new group of elements operating on functions, the element corresponding to A_i is denoted by P_{A_i} and is defined by the relation:

$$P_{A_i} f(A_i\mathbf{x}) = f(\mathbf{x}) \Leftrightarrow P_{A_i} f(\mathbf{x}) = f(A_i^{-1}\mathbf{x}) \tag{2.38}$$

These equations should hold for all $\mathbf{x}$ in the domain considered. The fact that the elements P_{A_i} defines a group can be derived directly by the group properties of $\mathcal{G}$ e.g:

$$\begin{aligned} &P_{A_i}P_{A_j} f(\mathbf{x}) = P_{A_i} f(A_j^{-1}\mathbf{x}) = f(A_j^{-1}A_i^{-1}\mathbf{x}) = f((A_iA_j)^{-1}\mathbf{x}) = P_{A_iA_j} f(\mathbf{x}) \\ &\Leftrightarrow P_{A_i}P_{A_j} = P_{A_iA_j} \end{aligned} \tag{2.39}$$

this shows that the new group satisfies the same multiplication table, and we say that it is *isomorphic* to the original group.

By a set of *basis functions* $\{f_1, ..., f_n\}$ for an n-dimensional representation Γ, we mean a set of functions satisfying the relation:

$$P_{A_i} f_\alpha = \sum_\beta \Gamma(A_i)_{\beta\alpha} f_\beta. \tag{2.40}$$

We also say that the functions $\{f_1, ..., f_n\}$ *belong* to this representation, or that they transform according to it.

2.6.1 Group of the Schrödinger equation.

Let us now consider a physical system with a symmetry group of coordinate transformations that leave the Hamiltonian invariant. This group is called the group of the Schrödinger equation. Then the group of operators P_{A_i} will all commute with the Hamiltonian. Then the time invariant Schrödinger equation gives:

$$\begin{aligned} H\psi_n &= E_n\psi_n \\ \Leftrightarrow P_{A_i}H\psi_n &= HP_{A_i}\psi_n = P_{A_i}E_n\psi_n \end{aligned} \tag{2.41}$$

This shows that $P_{A_i}\psi_n$ is also an eigenstate with the same energy E_n. Thus if we have one eigenstate we can generate other eigenstates with the same energy by applying all symmetry

operations, but they will not necessarily be independent, thus we do not know beforehand how many states to expect at a certain energy E_n. This procedure will not always generate all possible degenerate eigenstates with energy E_n, but when it does the degeneracy is said to be *normal*, otherwise it is said to be an *accidental* degeneracy. Accidental generacies are not originated by a symmetry. Often they turn out not to be exact or to have some origin of symmetry, when studied closer. Therefore, we will from here on assume no accidental degeneracies. Now consider an l_n-fold degenerate energy level of eigenfunctions $\psi_1^n, ..., \psi_{l_n}^n$ with energy E_n. By the result above any transformation P_{A_i} on any of the functions ψ_j^n will result in an eigenstate with the same energy, and thus can be written as a linear combination of these l_n functions. Thus the effect of the symmetry operations can be represented by matrix multiplication in this subspace:

$$P_{A_i}\psi_j^n = \sum_k \Gamma(A_i)_{kj}\psi_k^n \tag{2.42}$$

where the sum is over the degenerate eigenfunctions. The notation here is hinting that these matrices actually form a representation of the group, and in fact this turns out to be true:

$$\begin{aligned}\sum_m \Gamma(A_iA_l)_{mj}\psi_m^n = P_{A_iA_l}\psi_j^n = P_{A_i}P_{A_l}\psi_j^n = P_{A_i}\sum_k \Gamma(A_l)_{kj}\psi_k^n = \sum_{k,m}\Gamma(A_l)_{kj}\Gamma(A_i)_{mk}\psi_m^n \\ = \sum_m [\Gamma(A_i)\Gamma(A_l)]_{mj}\psi_m^n \end{aligned} \tag{2.43}$$

Then $\Gamma(A_iA_l) = \Gamma(A_i)\Gamma(A_l)$. Hence these Γ matrices form a representation, and the set of degenerate eigenfunctions are basis functions for this representations. It is in fact an irreducible representation. If it were reducible the basis for this degenerate subspace could then be transformed such that all the matrices of the representation became block diagonal, but this contradicts the fact that we assumed no accidental degeneracies. If the matrices were block diagonal then all the degenerate states $\psi_1^n, ..., \psi_{l_n}^n$ could not be generated by symmetry transformations of one particular eigenstate ψ_j^n which is the definition of an accidental degeneracy. Thus given the group of the Schrödinger equation the dimensionalities of the irreducible representations determine the possible non-accidental degeneracies of the system. These degeneracies can only be split by perturbations breaking the symmetry of the system. This shows how transformation properties can be used to label eigenstates, ψ_i^n is then the i'th eigenstate belonging to the n'th representation. The physical meaning of the first label depends on the basis chosen in the degenerate subspace.

2.7 Full rotation group and angular momentum.

In this section we will discuss the relation between the full rotatin group and angular momentum. We will assume that the reader is familiar with angular momentum in quantum mechanics, and relate this to the full rotation group.

The full rotation group consists of rotations by any angle around any axis going through the origin. This group is the first example, we have of an infinite group. It has an infinite number of classes, because any rotation of a certain angle α around any axis belong to the

same class, since the different rotation axes are related by another rotation. Therefore the number of representations is also infinite. The angular momentum operator **J** can be defined as the generator of rotations, which means that the operator P_{θ_z}, denoting a rotation of the angle θ_z around the z axis, can be written as $e^{-i\theta_z J_z}$. We can chose a basis of states with definite angular momentum in the z direction $|J,m\rangle$, where quantum numbers not concerning the angular momentum is left out. Since the total angular momentum is conserved, a rotation of $|J,m\rangle$ will be a superposition of states with the same J. Hence, the states $|J,m\rangle$ span a $2J+1$ dimensional representation of the full rotation group. Now we will calculate the characters of such a representation. Since any rotation axis can be transformed into any other rotation axis by a rotation, all rotations of the same angle will belong to the same class. So we can simply calculate the character of rotations around one specific axis. This is easiest done for the z axis:

$$P_{\theta_z}|J,m\rangle = e^{-im\theta_z}|J,m\rangle \tag{2.44}$$

and thus

$$\Gamma^J = \begin{pmatrix} e^{-iJ\theta_z} & 0 & ... & 0 \\ 0 & e^{-i(J-1)\theta_z} & ... & 0 \\ \vdots & & & \\ 0 & & & e^{iJ\theta_z} \end{pmatrix} \tag{2.45}$$

and we get the character:

$$\chi^J(P_{\theta_z}) = \sum_{m=-J}^{J} e^{-im\theta_z} = \frac{\sin((J+\frac{1}{2})\theta_z)}{\sin(\frac{\theta_z}{2})} \tag{2.46}$$

Since the character only depends on the angle, we will just use the notation $\chi^J(\theta)$. It can be shown that these Γ^J representations are all the irreducible representations of the full rotation group [8]. The character of a rotation of $\theta + 2\pi$ is given by:

$$\chi^J(\theta+2\pi) = \frac{\sin((J+\frac{1}{2})\theta + 2\pi(J+\frac{1}{2}))}{\sin(\frac{\theta}{2}+\pi)} = (-1)^{2J}\chi^J(\theta) \tag{2.47}$$

Thus if J is an half integer, then the character changes sign under rotations by 2π. This is in agreement with the fact that fermionic wavefunctions changes sign under 2π rotations [3]. To handle this we introduce a new symmetry operation $\mathcal{R}$ which is a rotation by 2π, but it is not equal to the identity operator since it changes the sign on fermionic states. This doubles the number of elements in the group, and therefore this new group is denoted the *double crystal group*. To construct the double group one only needs four rules, which we will simply state here. For a more complete discussion see [10].

1. If $\mathcal{C} = \{A_1, ..., A_n\}$ is a class of the original group then $\mathcal{C}$ and $\mathcal{RC} = \{\mathcal{R}A_1, ..., \mathcal{R}A_n\}$ both form classes in the double group. The only exception is if $\mathcal{C}$ is a class of rotations

[3] This fact was experimentally verified in 1975, by measuring the phase shift of neutrons precessing in a magnetic field [9].

by π around some axis and the symmetry group also contains rotations by π around an axis perpendicular to the first one. In that case $\mathcal{C}_2$ and $\mathcal{RC}_2$ belong to the same class.

2. Any irreducible representation of the original group is also an irreducible representation of the double group with the same set of characters $\chi(\mathcal{C})$, where C is any class in the original group, and $\chi(\mathcal{RC}) = \chi(\mathcal{C})$ for the new classes.

3. Since the number of irreducible representations is equal to the number of classes we must have some new irreducible representations. These new representations satisfies the relation $\chi(\mathcal{RC}) = -\chi(\mathcal{C})$.

2.7.1 Double group of D_{3d}

Now we want to double the group D_{3d}, using the above rules. Since inversion commutes with all elements of the double group, we can start by constructing the double group of D_3 and then take the direct product with the group of the inversion. We only have one class of rotations by π the number of classes is doubled and hence so is the number of irreducible representations. We have that $\sum_i l_i^2 = 12$ since we now have 24 elements in the group. If we sum only over the old representations then $\sum_i l_i^2 = 6$, and hence the sum over the 3 new representations is also $\sum_i l_i^2 = 6$. The only integer solution to this equation is if we have two new representations of dimension 1 and one of dimension 2. The new representations must give a minus sign when rotated by 2π, hence $\chi(\mathcal{C}_2)^2 = -1 \Leftrightarrow \chi(\mathcal{C}_2) = \pm i$ for the one dimensional representations. By the same argument $\chi(\mathcal{C}_3) = -1$ for the one dimensional representations. Now we have the character table except for two elements a and b, see table 2.6

	$\{E\}$	$2\mathcal{C}_3$	$3\mathcal{C}_2$	$\mathcal{R}$	$2\mathcal{RC}_3$	$3\mathcal{RC}_2$
$\Gamma^{(1)}$	1	1	1	1	1	1
$\Gamma^{(2)}$	1	1	-1	1	1	-1
$\Gamma^{(3)}$	2	-1	0	2	-1	0
$\Gamma^{(4)}$	1	-1	i	-1	1	-i
$\Gamma^{(5)}$	1	-1	-i	-1	1	i
$\Gamma^{(6)}$	2	a	b	-2	-a	-b

Table 2.6: Character table for the double group of D_3

a and b are easily found from the orthogonality relation for the columns:

$$-1 - 1 + 2a = 0 \Leftrightarrow a = 1 \tag{2.48}$$

$$2b = 0 \Leftrightarrow b = 0 \tag{2.49}$$

At last we can construct the character table for the double group of D_{3d} by taking the direct product with the group of inversion and the double group of D_3, giving the character table 2.7.

	$\{E\}$	$2\mathcal{C}_3$	$3\mathcal{C}_2$	I	$2I\mathcal{C}_3$	$3I\mathcal{C}_2$	$\mathcal{R}$	$2\mathcal{RC}_3$	$3\mathcal{RC}_2$	$\mathcal{R}I$	$2\mathcal{R}I\mathcal{C}_3$	$3\mathcal{R}I\mathcal{C}_2$
$\Gamma^{(1+)}$	1	1	1	1	1	1	1	1	1	1	1	1
$\Gamma^{(2+)}$	1	1	-1	1	1	-1	1	1	-1	1	1	-1
$\Gamma^{(3+)}$	2	-1	0	2	-1	0	2	-1	0	2	-1	0
$\Gamma^{(4+)}$	1	-1	i	1	-1	i	-1	1	$-i$	-1	1	$-i$
$\Gamma^{(5+)}$	1	-1	$-i$	1	-1	$-i$	-1	1	i	-1	1	i
$\Gamma^{(6+)}$	2	1	0	2	1	0	-2	-1	0	-2	-1	0
$\Gamma^{(1-)}$	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1
$\Gamma^{(2-)}$	1	1	-1	-1	-1	1	1	1	-1	-1	-1	1
$\Gamma^{(3-)}$	2	-1	0	-2	1	0	2	-1	0	-2	1	0
$\Gamma^{(4-)}$	1	-1	i	-1	1	$-i$	-1	1	$-i$	1	-1	i
$\Gamma^{(5-)}$	1	-1	$-i$	-1	1	i	-1	1	i	1	-1	$-i$
$\Gamma^{(6-)}$	2	1	0	-2	-1	0	-2	-1	0	2	1	0

Table 2.7: Character table for the double group of D_{3d}

2.8 Bloch's theorem

An important application of group theory in condensed matter physics is bloch's theorem. It applies to a physical system with translational symmetry. Let us just consider the one-dimensional case, generalization to higher dimensions is straightforward. If we have a periodic potential $V(x)$ of period a then the Hamiltonian:

$$H = \frac{p^2}{2m} + V(x) \tag{2.50}$$

is invariant under translations of an integer times a. Let this transformation be denoted by P_A, $P_A f(x) = f(x + a)$. For a finite system of length L we must have periodic boundary conditions to maintain translational invariance. Then we have a cyclic group of order $n = \frac{L}{a}$ since any element can be written A^m and $A^n = E$ because of the periodic boundary condition. Since this group is abelian we only have one dimensional representations. One dimensional representations are simply equal to their characters and since $A^n = E$ then the characters of any representation must be equal to $e^{im2\pi/n}$, where m is an integer which labels the representation. This gives exactly n distinct representations, since $m = m_0$ and $m = m_0 + n$ gives the same characters, and therefore the same representation. We can also label the representations by $k = \frac{2\pi m}{L}$. As we have seen any eigenfunction of the group of the Schrödinger equation must transform according to one of the irreducible representations of the group, we can label the eigenfunctions by k and ψ_k must satisfy the transformation rule:

$$P_A \psi_k(x) = \psi_k(x + a) = e^{ika} \psi_k(x) \tag{2.51}$$

This results in the well-known fact that eigenfunctions of a periodic potential can be written as a product of a free particle wavefunction times a function with the same periodicity as the potential:

$$\psi_k = e^{ikx} u_k(x) \tag{2.52}$$

where $u_k(x) = u_k(x + a)$.

2.9 Time reversal

A very special and important symmetry is the time reversal operator. The time reversal operator simply takes $t \to -t$. Classically this reverses all momenta, but leaves the position invariant. If we first consider the Schrödinger equation without spin:

$$(\frac{\nabla^2}{2m} + V(x))\psi(x,t) = i\frac{\partial\psi(x,t)}{\partial t} \tag{2.53}$$

Now we want to see how time reversal changes this equation, and we simply take $t \to -t$, which gives:

$$\begin{aligned}(\frac{\nabla^2}{2m} + V(x))\psi(x,-t) &= -i\frac{\partial\psi(x,-t)}{\partial t} \\ \Leftrightarrow (\frac{\nabla^2}{2m} + V(x))\psi^*(x,-t) &= i\frac{\partial\psi^*(x,-t)}{\partial t}\end{aligned} \tag{2.54}$$

We see that by complex conjugation we get the original Schrödinger equation back, but for $\psi^*(x,-t)$ instead of $\psi(x,t)$. Hence, if $\psi(x,t)$ is a solution so is $\psi^*(x,-t)$. This shows that for spinless particles the time reversal operator is simply comlex conjugation, denoted by K. It is very important to note that this operator, in contrast to almost any operator used in quantum mechanics is antilinear and not linear. An antilinear operator is defined by:

$$T(a|\psi\rangle + b|\varphi\rangle) = a^*T|\psi\rangle + b^*T|\varphi\rangle \tag{2.55}$$

where the difference from a linear operator is the complex conjugation. Furthermore this operator is *antiunitary*, which means that:

$$\langle T\psi|T\varphi\rangle = \langle\psi|\varphi\rangle^* \tag{2.56}$$

Again the difference from a unitary operator is the complex conjugation. This obviously holds for the complex conjugation operator:

$$\int(K\varphi)^*(K\psi)\,\mathrm{d}\mathbf{x} = \left(\int\varphi^*\psi\,\mathrm{d}\mathbf{x}\right)^* \tag{2.57}$$

For particles with spin, time reversal can also act in spin space, and since spin is angular momentum we expect time reversal to reverse the spin. Under time reversal the time evolution operator e^{iHt} must become $Te^{iHt}T^{-1} = e^{-iHt}$, for any particle regardless of spin and thus the time reversal operator is antilinear. And since we want it to conserve the norm it must be antiunitary. The product of two antiunitary operators must be unitary, and thus we have:

$$KT = U \Leftrightarrow T = UK \tag{2.58}$$

where U is some unitary operator. This actually means that any antiunitary operator can be

written as a product of complex conjugation operator and a unitary operator. Applying the time reversal operator twice should not change the physical state of the system which means that T^2 must be a trivial phase factor $e^{i\theta}$.

$$e^{i\theta} = T^2 = UKUK = UU^* \tag{2.59}$$

Since U is unitary $UU^\dagger = 1 \Leftrightarrow U^*U^T = 1$, so multiplying the last equality by U^T from the right gives:

$$U = e^{i\theta}U^T \Leftrightarrow U^T = Ue^{i\theta} \tag{2.60}$$

Substituting the second into the first gives:

$$U = e^{i\theta}Ue^{i\theta} = (e^{i\theta})^2U \Rightarrow (e^{i\theta})^2 = 1 \tag{2.61}$$

This eqaution gives that $T^2 = e^{i\theta} = \pm 1$. Using $T^2 = UU^*$ we see that the determinant of T^2 is:

$$\det(T^2) = \det(UU^*) = \det(U)\det(U^*) = \det(U)\det(U)^* = 1 \tag{2.62}$$

But $T^2 = \pm 1$ which means that the determinant is $(\pm 1)^n$ where n is the dimension of the space U is operating in. This is the spin space, which is odd dimensional for integer spin. Hence, $T^2 = 1$ for integer spin.

If we have a half-integer spin instead, T^2 is not determined from this consideration. From a classical point of view we expect the angular momentum to be reversed under time reversal. This means that $T\mathbf{J}T = -\mathbf{J}$ for any angular momentum operator J is the spin operator. For a spin $\frac{1}{2}$ system we have in general that $T = UK$. We choose a basis where the spinoperators are represented by the pauli matrices $\mathbf{S} = \frac{1}{2}\sigma$, with:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \tag{2.63}$$

Any 2 by 2 matrix can be written as a linear combination of the pauli matrices and the identity matrix, so if we write:

$$U = \mathbf{d} \cdot \sigma + d_0\sigma_0, \tag{2.64}$$

where $\sigma = (\sigma_x, \sigma_y, \sigma_z)$ and σ_0 is the identity matrix.

The requirement that T reverses spin then results in the following:

$$T\sigma_xT^{-1} = U\sigma_xU^{-1} = -\sigma_x \Leftrightarrow U\sigma_x = -\sigma_xU \tag{2.65}$$

$$T\sigma_yT^{-1} = -U\sigma_yU^{-1} = -\sigma_y \Leftrightarrow U\sigma_y = \sigma_yU \tag{2.66}$$

$$T\sigma_zT^{-1} = U\sigma_zU^{-1} = -\sigma_z \Leftrightarrow U\sigma_z = -\sigma_zU \tag{2.67}$$

U must commute with σ_y and anticommute with σ_x and σ_z. The first equation, gives that $d_1 = d_0 = 0$ and the third that $d_3 = d_0 = 0$ and hence $U = d_2\sigma_y$. Since U is unitary d_2 is just a phase factor. Then we have:

$$T^2 = d_2\sigma_y K d_2\sigma_y K = -|d_2|^2\sigma_y^2 = -1 \tag{2.68}$$

Usually the phase is chosen such that $T = i\sigma_y K$. This shows that $T^2 = -1$ for a spin half system, but it actually turns out to be the case for any non-integer spin system.

2.9.1 Kramers Theorem

For a system of non-integer spin time reversal invariance leads to a double degeneracy, known as Kramers degeneracy. This degeneracy is a direct result of the fact that $T^2 = -1$. If the system is described by the Hamiltonian H is time reversal invariant then T and H commutes. If we have an eigenstate $|\psi\rangle$ with energy E, then:

$$H|\psi\rangle = E|\psi\rangle \Leftrightarrow TH|\psi\rangle = HT|\psi\rangle = ET|\psi\rangle \tag{2.69}$$

and $T|\psi\rangle$ is also an eigenstate of energy E. However this does not tell us if we have a degeneracy, since $|\psi\rangle$ could be proportional to $|\psi\rangle$, but if $T^2 = -1$ then:

$$\begin{aligned} &\langle\psi|T\psi\rangle = -\langle T^2\psi|T\psi\rangle = -\langle T\psi|\psi\rangle^* = -\langle\psi|T\psi\rangle \\ \Leftrightarrow\ &\langle\psi|T\psi\rangle = 0 \end{aligned} \tag{2.70}$$

showing that $|\psi\rangle$ and $T|\psi\rangle$ are orthogonal, and we have a double degeneracy. Note that the only proporties of T used in this derivation is $T^2 = -1$ and the fact that T is antiunitary.

2.9.2 Inversion and time reversal

Now we turn to a translational invariant system again. The eigenstates of the Hamiltonian can then be written as bloch states which under lattice translations transforms as:

$$\psi_{n,\mathbf{k}}(\mathbf{x} + \mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}}\psi_{n,\mathbf{k}}(\mathbf{x}) \tag{2.71}$$

Then the time reversed of this state transforms as:

$$T\psi_{n,\mathbf{k}}(\mathbf{x} + \mathbf{R}) = e^{-i\mathbf{k}\cdot\mathbf{R}}T\psi_{n,\mathbf{k}}(\mathbf{x}) \tag{2.72}$$

then $T\psi_{n,k}$ must be a state of opposite crystal momentum $\psi_{m,-k}$. Note that the label m has changed, and that it also contains the spin. This shows that for a time reversal invariant system the dispersion relation $E_n(\mathbf{k})$ must be symmetric around $\mathbf{k} = 0$, and specifically at $\mathbf{k} = 0$ all bands have at least a double degeneracy.

For a system with both inversion I and time reversal symmetry then we can use that the system also has the symmetry IT. This an antiunitary operator since I is unitary and $(IT)^2 = -1$, and hence the states related by this symmetry must be orthogonal because of

Kramers theorem. The state $IT\psi_{n,\mathbf{k}}$ transforms under lattice translations as:

$$IT\psi_{n,\mathbf{k}}(\mathbf{x}+\mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}} IT\psi_{n,\mathbf{k}}(\mathbf{x}) \tag{2.73}$$

since inversion takes $\mathbf{x} \to -\mathbf{x}$. This means that the states related by IT are at the same $\mathbf{k}$ value, and all bands of a system with both inversion and TR symmetry are doubly degenerate.

2.10 Symmetries of crystals

In solid state physics, we consider crystals, regular arrays of identical unit cells consisting of one or more atoms. These will of course have some symmetries. The group of all coordinate transformations leaving a crystal invariant is called the *space group* of the crystal. A subgroup of the space group, is the group of lattice translations by a vector $\mathbf{T}_n = n_1\mathbf{a}_1 + n_2\mathbf{a}_2 + n_3\mathbf{a}_3$, where n_i are integers and $\mathbf{a}_i$ are the primitive translation vectors, defining the unit cell. In addition there are transformations $\mathbf{x} \to \mathbf{R}\mathbf{x}$ where $\mathbf{R}$ is an orthogonal matrix. These are rotations, inversion and combinations of those (or proper and improper rotations). A general element is usually written $\{\mathbf{R}|\mathbf{T}\}$. The group we get by putting $T = 0$ is called the *point group*. If the all elements of the point group are symmetries of the crystal, the space group is said to be *symmorphic*. In a non-symmorphic group there are elements which involves both a translation and a rotation, each of which are not symmetries of the crystal. But here we will only consider the symmorphic groups. The three dimensional version of bloch's theorem states that the eigenfunctions of a translationally invariant system can be written:

$$\psi_{\mathbf{k}}(\mathbf{x}) = u_{\mathbf{k}}(\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{x}} \tag{2.74}$$

The effect of a transformation $\mathbf{R}$ from the point group on the wavefunction is then:

$$P_{\mathbf{R}}\psi_{\mathbf{k}}(\mathbf{x}) = u_{\mathbf{k}}(\mathbf{R}^{-1}\mathbf{x})e^{i\mathbf{k}\cdot\mathbf{R}^{-1}\mathbf{x}} = u_{\mathbf{k}}(\mathbf{R}^{-1}\mathbf{x})e^{i\mathbf{R}\mathbf{k}\cdot\mathbf{x}} \tag{2.75}$$

Here we used the orthogonality of the matrix $\mathbf{R}$, $\mathbf{R}^{-1} = \mathbf{R}^T$. $u_{\mathbf{k}}(\mathbf{R}^{-1}\mathbf{x})$ is also periodic with the periodicity of the lattice, since $\mathbf{R}$ is a symmetry of the lattice. Thus applying the operator $P_{\mathbf{R}}$ on a bloch wave function, gives a bloch wave function at crystal momentum $\mathbf{R}\mathbf{k}$.

2.11 Theory of invariants

In this section we will describe how to construct invariant linear combinations of objects belonging to irreducible representations of a symmetry group. To do this we will need Schur's lemma part 1 & 2, which we will just state here, for a proof se [10].

Schur's lemma Part 1 A matrix which commutes with all matrices of an irreducible representation is a constant times the identity matrix. Therefore, if a non-constant commuting matrix exists, the representation is reducible.

Schur's lemma Part 2 If $\Gamma^{(1)}$ and $\Gamma^{(2)}$ are two irreducible representations of a given

group $\mathcal{G}$ of dimensionality l_1 and l_2, then if a $l_1 \times l_2$ matrix exists such that:

$$\Gamma^{(1)}(X)M = M\Gamma^{(2)}(X) \quad \forall X \in \mathcal{G} \tag{2.76}$$

Then if $l_1 \neq l_2$ then M must be the null matrix ($M = 0$), and if $l_1 = l_2$ then either $M = 0$ or $\Gamma^{(1)}$ and $\Gamma^{(2)}$ are equivalent.

Consider a symmetry group $\mathcal{G}$, and two sets of objects $\mathbf{u} = (u_1, ..., u_n)$ and $\mathbf{v} = (v_1, .., v_m)$ transforming according to the irreducible representations $\Gamma^{(u)}$ and $\Gamma^{(v)}$, which are n and m dimensional. We also choose them to be unitary [4]. Then a linear combination $\mathbf{u}^\dagger M v$ (where M is a $n \times m$ matrix), which is invariant can be constructed if and only if $\Gamma^{(u)}$ and $\Gamma^{(v)}$ are equivalent.

Proof. First, if $\Gamma^{(u)}$ and $\Gamma^{(v)}$ are equivalent then there exists some matrix A such that $\Gamma^{(u)}(X) = A^{-1}\Gamma^{(v)}(X)A$ for all $X \in \mathcal{G}$. Then under the transformation X we have:

$$\mathbf{u}^\dagger M\mathbf{v} \to \mathbf{u}^\dagger\Gamma^{(u)}(X)^\dagger M A^{-1}\Gamma^{(u)}(X)A\mathbf{v} \tag{2.77}$$

We can then choose $M = A$ and then we have an invariant term.

And second, if $\mathbf{u}^\dagger M\mathbf{v}$ is invariant, then for all $X \in \mathcal{G}$:

$$\mathbf{u}^\dagger M\mathbf{v} = \mathbf{u}^\dagger(\Gamma^{(u)})^\dagger(X)M(\Gamma^{(v)})^\dagger(X)\mathbf{v} \tag{2.78}$$

$$\Leftrightarrow M = (\Gamma^{(u)})^\dagger(X)M(\Gamma^{(v)})^\dagger(X) \tag{2.79}$$

Then either $M = 0$ or $\Gamma^{(u)}$ and $\Gamma^{(v)}$ are equivalent by Schur's lemma part 2. Then to get a nontrivial invariant term $\Gamma^{(u)}$ and $\Gamma^{(v)}$ must be equivalent. □

Hence all invariant terms we can construct must be combinations of objects from the same irreducible representation.

[4] Any representation by matrices is equivalent to a representation by unitary matrices [8]

Low energy effective model

In this chapter, we will derive the form of the low energy effective model for Bi_2Se_3. We want to find a simple model that describes the topological nature of this material. Since we have both time reversal and inversion symmetry all bands are doubly degenerate and we need at least a four band model to get a gapped spectrum. That is our goal for this chapter; to obtain a four band model that describes the low energy physics of Bi_2Se_3. We will use the theory of invariants, from section 2.11 to write down the most general model allowed by the symmetries of the crystal. When this is done, the parameters can be found by fitting the spectrum to experimental data or *ab initio* calculations of the spectrum. Another method is to use what is known as $\mathbf{k} \cdot \mathbf{p}$ perturbation theory, where the parameters can be determined from matrix elements of the momentum operator with wave functions at the gamma point ($\mathbf{k} = 0$) from *ab initio* calculations. Here we will only construct the model, and we will use both the parameters of [11] and [12]. Since we only use the symmetries of the crystal this holds for any material of the same structure, e.g. Bi_2Te_3. This class of materials is denoted the tetradymite group.

3.1 Crystal structure

In this section, we will describe the crystal structure of Bi_2Se_3. Understanding the structure is essential, when figuring out the symmetries of the crystal and considering which basis states to use for our effective model. The crystal structure of Bi_2Se_3 is shown in figure 3.1. It

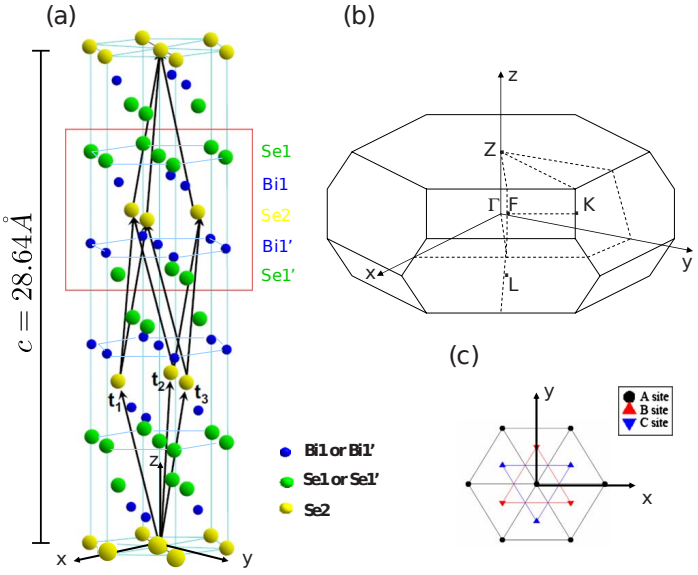


Figure 3.1: (a) The crystal structure of Bi_2Se_3, with one quintuple layer framed by the red box. The positions the triangular lattice of each layer can be A, B or C sites as indicated on (c). The positions of the layers alternate A-C-B-A-C-B etc. while the types of atoms are arranged in quintuple layers consisting of five atomic layers in the order Se-Bi-Se-Bi-Se. (b) shows the first Brillouin zone of the lattice. Adapted from [11].

is a layered structure, with each layer consisting of a triangular lattice of either bismuth or

selenium atoms. We choose a coordinate system with the z-axis perpendicular to the atomic layers. The positions of the layers alternate between the A, B and C sites. The structure can be divided into the so-called quintuple layers consisting of five atomic layers, alternating between bismuth layers and selenium layers. The primitive translation vectors are given by:

$$\mathbf{t}_1 = \begin{pmatrix} \frac{a}{2} \\ -\frac{\sqrt{3}a}{6} \\ \frac{c}{3} \end{pmatrix}, \mathbf{t}_2 = \begin{pmatrix} -\frac{a}{2} \\ -\frac{\sqrt{3}a}{6} \\ \frac{c}{3} \end{pmatrix}, \mathbf{t}_3 = \begin{pmatrix} 0 \\ \frac{\sqrt{3}a}{3} \\ \frac{c}{3} \end{pmatrix}, \tag{3.1}$$

where $a = 4.138\,\text{Å}$ is the lattice constant of the layered triangular lattices, and $c = 28.64\,\text{Å}$ is the length of the longest diagonal of the rhombohedral unit cell. On figure 3.1 it is the distance between the top and bottom layers. The values of a and c are from [13]. Then the thickness of one quintuple layer is $\frac{c}{3} = 9.547\,\text{Å}$. There are five atoms in each primitive unit cell, two equivalent bismuth atoms denoted Bi1 and Bi1′, two equivalent selenium atoms Se1 and Se1′ and a selenium atom which is not equivalent to the two others. By equivalent we mean that their positions can be interchanged by a symmetry operation of the crystal. One rhombohedral unit cell is indicated by the primitive translation vectors shown on figure 3.1, and the 5 atoms in this unit cell are the ones exactly at the z-axis. If we take the origin to be in a Se2 atom then the positions of the other atoms in the rhombohedral unit cell are given by $\pm 0.399c\hat{z}$ for Bi1 and Bi1′ and $\pm 0.206c\hat{z}$ for Se1 and Se1′, according to [13]. Note that these atoms are in different quintuple layers. The distances between the atomic layers within one quintuple layer can be calculated using that the thickness of one quintuple layer is $\frac{c}{3}$. The distance between the Se2 layer and the Bi1 or Bi1′ layers is $0.399c - \frac{c}{3} = 0.066c = 1.890\,\text{Å}$, the distance between Bi1(′) and Se1(′) layers $-0.206c + \frac{c}{3} - 0.066c = 0.061c = 1.747\,\text{Å}$. The distance between a Se1 and the Se1′ of the neighbouring quintuple layer is $0.206c - (-0.206c + \frac{c}{3}) = 0.079c = 2.263\,\text{Å}$. The vector from one atom to one of the nearest atoms in the next layer above is $\mathbf{n} = (0, -\frac{\sqrt{3}a}{3}, h)$ where h is the distance between the layers. Then the distance between the two atoms is $\sqrt{\frac{a^2}{3} + h^2}$. The three vectors connecting one atom to the three nearest in the next layer are related to each other by rotations of 120 degrees, and the angle between two of these is given by:

$$\cos(\theta) = \frac{\mathbf{n} \cdot R_3\mathbf{n}}{\mathbf{n} \cdot \mathbf{n}} = \frac{\frac{a^2}{3}\cos(120°) + h^2}{\frac{a^2}{3} + h^2} = \frac{-\frac{a^2}{6} + h^2}{\frac{a^2}{3} + h^2}, \tag{3.2}$$

where R_3 is a matrix, representing a rotation of 120° around the z axis. We denote the angle between the two vectors connecting a Se2 atom and two of the nearest neighbours in the Bi1 layer α, and the angle between the two vectors connecting a Bi1 atom to two of the nearest neighbours in the Se1 layer β. These angles are shown in figure 3.2 and using eq. (3.2) we get:

$$\alpha = 85.6° \quad \beta = 88.7° \tag{3.3}$$

We see that the vectors connecting one atom to the three nearest atoms in the next layer, but still in the same quintuple layer, are almost orthogonal.

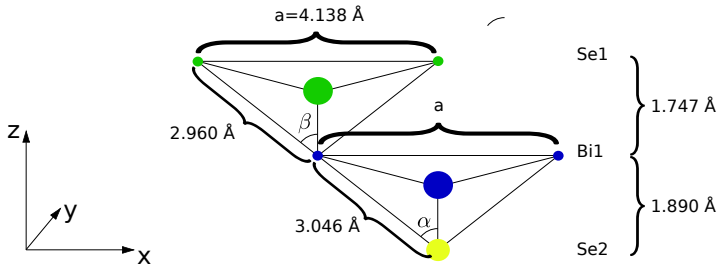


Figure 3.2: The relative positions of atoms in different layers in 3 dimensions. The color indicates the type of atom, and all atoms of the same type are in the same layer, i.e. has the same position on the z axis. The size of the atoms indicate the position on the y axis. The positions of Bi1' and Se1' can be found by inversion of the Bi1 and Se1. Note that the nearest neighbours are actually from different layers. The angles shown are $\alpha \approx \beta \approx 90°$, which means that the 3 p orbitals in the direction of the nearest neighbours are approximately orthorgonal.

3.1.1 Symmetry group

The space group of this crystal consists of the group of translations by a vector:

$$R = n_1\mathbf{t}_1 + n_2\mathbf{t}_2 + n_3\mathbf{t}_3 \quad n_1, n_2, n_3 \in \mathbb{Z} \tag{3.4}$$

and the point group D_{3d}. The point group D_{3d} is the direct product group of D_3 and the group of the inversion operator. The effect of these symmetries are easiest to visualize by considering the rhombohedral unit cell and the three primitive translation vectors.

1. Rotation around the z-axis by 120 degrees. Since all the atoms in the unit cell lie on the z axis any rotations around the z-axis does not change anything within the unit cell. This rotation transforms $(\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3) \to (\mathbf{t}_3, \mathbf{t}_1, \mathbf{t}_2)$ and thus the lattice is invariant.

2. C_2 rotations around the x-axis. This rotation turns the unit cell upside down, but that simply changes interchanges Bi1 (Se1) and Bi1′ (Se1′). This rotation transforms $(\mathbf{t}_1, \mathbf{t}_2, \mathbf{t}_3) \to (-\mathbf{t}_2, -\mathbf{t}_1, \mathbf{t}_3)$ and thus the lattice is invariant.

3. Inversion I. This transformation takes $\mathbf{r} \to -\mathbf{r}$ and thus it also interchanges Bi1 (Se1) and Bi1′ (Se1′). Inversion transforms $\mathbf{t}_i \to -\mathbf{t}_i$ and again this leaves the lattice invariant.

All elements of the point group D_{3d} can be produced by combination of these three transformations.

3.2 Qualitative description of the basis states around the Fermi level

The goal of this section is to find the basis states for our effective model, i.e. the states closest around the Fermi level at the gamma point. Starting from the atomic orbitals of bismuth and selenide we will give qualitative explanations of the essential physics. Following [11], we will do this in three steps; first we consider the nearest neighbour coupling within one quintuple layer, then the crystal field splitting and at last the spin-orbit coupling. See figure 3.3 for a schematic picture of the results of this section. The arguments of this section follows [11], where the qualitative results are also supported by ab initio band structure calculations.

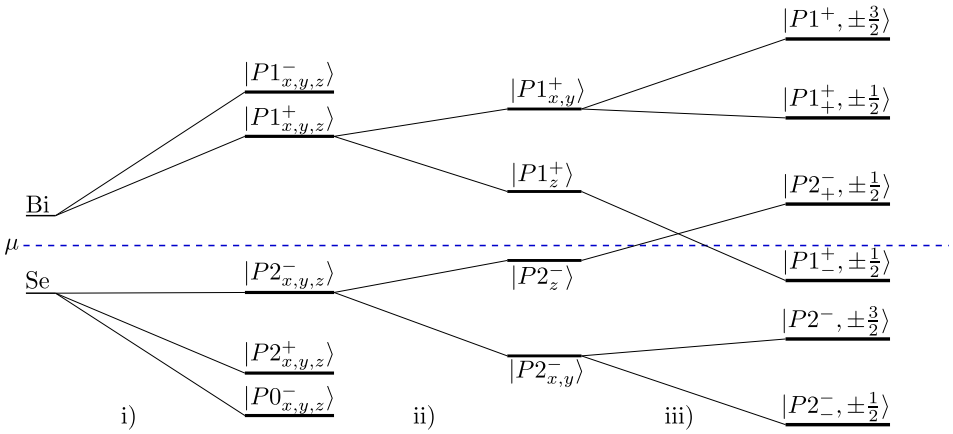


Figure 3.3: Qualitative picture of the splittings of the outermost atomic orbitals. The starting point is the $6p$ atomic level of bismuth and the $4p$ level of selenide. The degenaracies of these level are 12 for bismuth and 18 for selenide, since there is two bismuth atoms and three selenium atoms in each unit cell. Each atom has 3 different p orbitals and two different spin states. i) first we consider the splitting due to the nearest neighbour couplings within one quintuple layer, then ii) the crystal field splitting and iii) we see a crossing between the states closest to the Fermi level when we include the spin-orbit coupling.

3.2.1 Coupling within a quintuple layer

The electron configuration of bismuth is $6s^2 6p^3$ while selenium has $4s^2 4p^4$. Therefore, the outermost orbitals for both atoms are the p orbitals and we will neglect all other orbitals. This gives a total of 30 states in one unit cell, 3 p orbitals for each atom, and two different spin states. The number of electrons in one unit cell is 18, three for each of the two bismuth atoms and four for each of the three selenium atoms. Therefore the 18 lowest lying states will be filled, and the chemical potential is somewhere above these states, but below the other states. Hence, in the atomic limit (the starting point on the left side of figure 3.3) the chemical potential is between the two atomic levels. In [14] they find from density functional theory calculations, that the bismuth atoms are positively charged, while the selenium atoms are negatively charged. Furthermore, they argue that this indicates that the coupling within one quintuple layer is of the covalent-ionic type while the coupling between different layers are of the weaker van der Waals type. Thus, the simplest model one could imagine is simply a tight binding model within one quintuple layer. We have 3 p orbitals in the five different layers. Now in principle one has to consider the coupling between any p orbital of one layer with any p orbital of the next layer, but we can change to a basis where couplings between orbitals of the same type dominate. We can change from p_x, p_y and p_z orbitals to a basis of p orbitals pointing in the directions of the nearest neighbours in the next layer. Let these three be denoted by p_a, p_b and p_c, they are shown in figure 3.4. These are to a good approximation orthogonal as we saw above. Then there is almost no overlap between different types of p orbitals in the different layers, while there is a big overlap between p orbitals of the same type, see 3.4. This means that for each of these three we have the Hamiltonian:

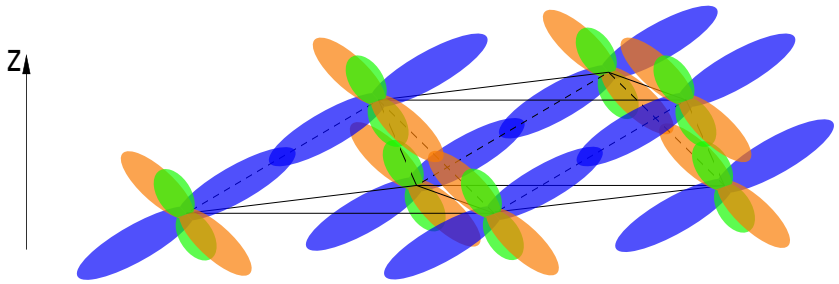


Figure 3.4: If the basis for the p orbitals are chosen in the directions of the nearest neighbours, then orbitals of one layer have large overlap with orbitals of the same type in the next layer.

$$H_{tb} = \begin{pmatrix} \varepsilon_{\mathrm{Se}} & t & 0 & 0 & 0 \\ t & \varepsilon_{\mathrm{Bi}} & t & 0 & 0 \\ 0 & t & \varepsilon_{\mathrm{Se}} & t & 0 \\ 0 & 0 & t & \varepsilon_{\mathrm{Bi}} & t \\ 0 & 0 & 0 & t & \varepsilon_{\mathrm{Se}} \end{pmatrix}, \tag{3.5}$$

in the basis of $|\mathrm{Se}1, p_\alpha\rangle, |\mathrm{Bi}1, p_\alpha\rangle, |\mathrm{Se}2, p_\alpha\rangle, |\mathrm{Bi}1', p_\alpha\rangle, |\mathrm{Se}1', p_\alpha\rangle$, where $\alpha \in \{a, b, c\}$. Here $\varepsilon_{\mathrm{Se}}$

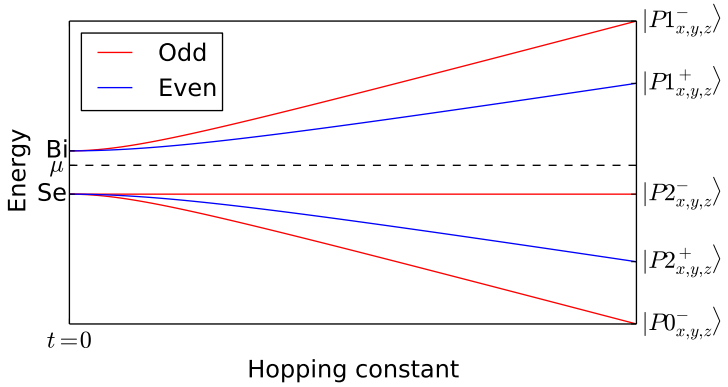


Figure 3.5: The energy levels of one quintuple layer, as a function of the hopping constant t. The five resulting levels consist of 6 degenerate states; three different p orbitals and two different spin states. The color shows the eigenvalue of the various states under inversion. Since the p_a, p_b, p_c orbitals are split in the same way, we have changed the basis back to p_x, p_y, p_z.

and $\varepsilon_{\mathrm{Bi}}$ are the atomic energies of bismuth and selenium. We use the same t for all couplings, since this is only a qualitative model and since the distance between the layers are almost the same. Turning on the hopping constant t splits the atomic levels as shown in figure 3.5. Since we have inversion symmetry, the levels are split eigenstates which are even or odd under inversion. The inversion eigenvalues are indicated by the superscript. Note that since the p orbitals are odd under inversion, and inversion interchanges the Bi1 (Se1) and Bi1′ (Se1′), the

inversion operator in this basis is given by:

$$I = \begin{pmatrix} 0 & 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \end{pmatrix}. \tag{3.6}$$

Since the three different p orbitals are split the same way, we can change our basis back to p_x, p_y and p_z. From now on, we focus on the two levels closest to the Fermi level, which is between the two atomic energies, $|P1^+_{x,y,z}\rangle$ and $|P2^-_{x,y,z}\rangle$.

3.2.2 Crystal field splitting

Next, we consider the crystal field splitting. The atomic orbitals have full rotation symmetry, but this symmetry can be broken by the crystal field. However, the crystal field is symmetric under D_3, which will determine how the degenaracies when the p orbitals are split[1]. The p orbitals are basis functions for an irreducible representation of the full rotation group, $\Gamma^{(J=1)}$. When the full rotation symmetry is broken down to D_3 this representation is no longer irreducible. The characters of the elements of D_3 are according to eq. (2.46):

$$\chi^{(J=1)}(E) = 3, \quad \chi^{(J=1)}(\mathcal{C}_2) = \frac{\sin(\frac{3}{2}\pi)}{\sin(\frac{\pi}{2})} = -1, \quad \chi^{(J=1)}(\mathcal{C}_3) = \frac{\sin(\pi)}{\sin(\frac{\pi}{3})} = 0, \tag{3.7}$$

The character of the identity is 3, since it is a 3 dimensional representation. By inspection of the character table of the group D_3 (table 2.3 p. 12) we see that $\Gamma^{(J=1)}$ is a reducible representation of D_3 which can be written in terms of the irreducible representations of D_3 as

$$\Gamma^{(J=1)} = \Gamma^{(2)} + \Gamma^{(3)} \tag{3.8}$$

With full rotation symmetry the three p orbitals are degenerate, but lowering the symmetry to D_3 splits this degeneracy into to a non-degenerate and a doubly degenerate level. It must be p_z that belongs to the one dimensional representation $\Gamma^{(2)}$ and p_x and p_y that belongs to $\Gamma^{(3)}$ and hence still have a degeneracy. The symmetry considerations here only shows that the three p orbitals can be split into two levels; the p_z orbital and a degenerate level of p_x and p_y. It turns out that energy of the $|P1^+_{x,y}\rangle$ are increased while $|P1^+_z\rangle$ decreases, and the other way around for the states with negative inversion eigenvalue, according to [11]. Thus, both $|P1^+_z\rangle$ and $|P2^-_z\rangle$ gets closer to the chemical potential.

3.2.3 Spin-orbit coupling

At this point all the states are doubly degenerate because of the spin. Now we will consider the effect of spin-orbit coupling on the atomic orbitals. The atomic spin-orbit Hamiltonian is

[1] Here we only consider the group D_3, which does not include inversion, since inversion symmetry does not change the degeneracies.

given by [15]:

$$H_{so} = \lambda \mathbf{s} \cdot \mathbf{L} = \frac{1}{2m_0^2 c^2 r} \frac{\partial U}{\partial r} \mathbf{s} \cdot \mathbf{L} \tag{3.9}$$

where U is the potential of the atoms, m_0 is the electron mass, c is the speed of light, and r is the position relative to the center of the nucleus. Now we consider the effect of this interaction on the states $|P1_\alpha^+, m_s\rangle$ and $|P2_\alpha^-, m_s\rangle$, where α denotes the type of p orbital p_x, p_y or p_z while m_s denotes the spin in the z direction $\pm\frac{1}{2}$, giving in total 12 states. Now it is convenient to change the basis to a basis with definite orbital angular momentum in the z direction. This basis can be written $|\Lambda, m_l, m_s\rangle$, where $\Lambda \in \{P1^+, P2^-\}$, m_l and m_s are the projections in the z direction of orbital and spin angular momenta. In terms of the old basis these states are given by:

$$|\Lambda, 1, m_s\rangle = -\frac{1}{\sqrt{2}}(|\Lambda_x, m_s\rangle + i|\Lambda_y, m_s\rangle) \tag{3.10}$$

$$|\Lambda, 0, m_s\rangle = |\Lambda_z, m_s\rangle \tag{3.11}$$

$$|\Lambda, -1, m_s\rangle = \frac{1}{\sqrt{2}}(|\Lambda_x, m_s\rangle - i|\Lambda_y, m_s\rangle) \tag{3.12}$$

Since the spin-orbit coupling is rotationally invariant (when rotation is applied in both spin and orbital space) total angular momentum is conserved. The spin-orbit term can be rewritten using the raising and lowering operators $L_\pm = L_x \pm iL_y$ and $S_\pm = S_x \pm iS_y$:

$$\mathbf{S} \cdot \mathbf{L} = S_x L_x + S_y L_y + S_z L_z = \frac{1}{2}(S_+ L_- + S_- L_+) + S_z L_z \tag{3.13}$$

the effect of this operator on our basis states is:

$$\mathbf{S} \cdot \mathbf{L}|\Lambda, 1, \tfrac{1}{2}\rangle = \frac{1}{2}|\Lambda, 1, \tfrac{1}{2}\rangle \tag{3.14}$$

$$\mathbf{S} \cdot \mathbf{L}|\Lambda, 0, \tfrac{1}{2}\rangle = \frac{1}{\sqrt{2}}|\Lambda, 1, -\tfrac{1}{2}\rangle \tag{3.15}$$

$$\mathbf{S} \cdot \mathbf{L}|\Lambda, -1, \tfrac{1}{2}\rangle = \frac{1}{\sqrt{2}}|\Lambda, 0, -\tfrac{1}{2}\rangle - \frac{1}{2}|\Lambda, -1, \tfrac{1}{2}\rangle \tag{3.16}$$

$$\mathbf{S} \cdot \mathbf{L}|\Lambda, 1, -\tfrac{1}{2}\rangle = \frac{1}{\sqrt{2}}|\Lambda, 0, \tfrac{1}{2}\rangle - \frac{1}{2}|\Lambda, 1, -\tfrac{1}{2}\rangle \tag{3.17}$$

$$\mathbf{S} \cdot \mathbf{L}|\Lambda, 0, -\tfrac{1}{2}\rangle = \frac{1}{\sqrt{2}}|\Lambda, -1, \tfrac{1}{2}\rangle \tag{3.18}$$

$$\mathbf{S} \cdot \mathbf{L}|\Lambda, -1, -\tfrac{1}{2}\rangle = \frac{1}{2}|\Lambda, -1, -\tfrac{1}{2}\rangle \tag{3.19}$$

As we expected the total angular momentum in the z direction $m_l + m_s$ is conserved. Therefore, the states $|\Lambda, \pm 1, \pm\frac{1}{2}\rangle$ are already eigenstates of the spin-orbit Hamiltonian. The non-zero

matrix elements coupling the other states are:

$$\langle \Lambda, 1, -\tfrac{1}{2}|H_{so}|\Lambda, 0, \tfrac{1}{2}\rangle = \langle \Lambda, 0, -\tfrac{1}{2}|H_{so}|\Lambda, -1, \tfrac{1}{2}\rangle = \frac{\lambda_\Lambda}{\sqrt{2}} \tag{3.20}$$

$$\langle \Lambda, 1, \tfrac{1}{2}|H_{so}|\Lambda, 1, \tfrac{1}{2}\rangle = \langle \Lambda, -1, -\tfrac{1}{2}|H_{so}|\Lambda, -1, -\tfrac{1}{2}\rangle = \frac{\lambda_\Lambda}{2} \tag{3.21}$$

$$\langle \Lambda, -1, \tfrac{1}{2}|H_{so}|\Lambda, -1, \tfrac{1}{2}\rangle = \langle \Lambda, 1, -\tfrac{1}{2}|H_{so}|\Lambda, 1, -\tfrac{1}{2}\rangle = -\frac{\lambda_\Lambda}{2} \tag{3.22}$$

Where we have introduced a spin-orbit constant λ_Λ for each band, since they could be different. Since the spin-orbit coupling only couples states with the same total angular momentum in the z direction m_j the Hamiltonian becomes block diagonal with blocks corresponding to each value of m_j. The spin-orbit coupling is also diagonal in the $P1^+$,$P2^-$ space, and each block can be written:

$$H_{\Lambda,so} = \begin{pmatrix} E_{\Lambda,x} + \frac{\lambda_\Lambda}{2} & 0 & 0 & 0 \\ 0 & h_{\Lambda,\frac{1}{2}} & \mathbf{0} & 0 \\ 0 & \mathbf{0} & h_{\Lambda,-\frac{1}{2}} & 0 \\ 0 & 0 & 0 & E_{\Lambda,x} + \frac{\lambda_\Lambda}{2} \end{pmatrix}, \tag{3.23}$$

in the basis $|\Lambda, 1, \frac{1}{2}\rangle, |\Lambda, 1, -\frac{1}{2}\rangle, |\Lambda, 0, \frac{1}{2}\rangle, |\Lambda, -1, \frac{1}{2}\rangle, |\Lambda, 0, -\frac{1}{2}\rangle, |\Lambda, -1, -\frac{1}{2}\rangle$. The spin $\frac{1}{2}$ blocks are given by:

$$h_{\Lambda,\frac{1}{2}} = h_{\Lambda,-\frac{1}{2}} = \begin{pmatrix} E_{\Lambda_x} - \frac{\lambda_\Lambda}{2} & \frac{\lambda_\Lambda}{\sqrt{2}} \\ \frac{\lambda_\Lambda}{\sqrt{2}} & E_{\Lambda_z} \end{pmatrix}, \tag{3.24}$$

where E_{Λ_α} is the energy of the $|\Lambda_\alpha\rangle$ without spin-orbit coupling. The remaining four eigenstates of the spin-orbit coupling can now be found just by solving the 2×2 Hamiltonian $h_{\Lambda,\pm\frac{1}{2}}$. The six eigenstates of the each Λ block of the spin-orbit coupling can be written:

$$|\Lambda, \pm\tfrac{3}{2}\rangle = |\Lambda, \pm 1, \pm\tfrac{1}{2}\rangle \quad E = E^\Lambda_{\frac{3}{2}} \tag{3.25}$$

$$|\Lambda_+, \pm\tfrac{1}{2}\rangle = u^\Lambda_+|\Lambda, \pm 1, \mp\tfrac{1}{2}\rangle + v^\Lambda_+|\Lambda, 0, \pm\tfrac{1}{2}\rangle \quad E = E^{\Lambda_+}_{\frac{1}{2}} \tag{3.26}$$

$$|\Lambda_-, \pm\tfrac{1}{2}\rangle = u^\Lambda_-|\Lambda, \pm 1, \mp\tfrac{1}{2}\rangle + v^\Lambda_-|\Lambda, 0, \pm\tfrac{1}{2}\rangle \quad E = E^{\Lambda_-}_{\frac{1}{2}} \tag{3.27}$$

where

$$u^\Lambda_\pm = \frac{E_{\Lambda,x} - E_{\Lambda,z} - \frac{\lambda_\Lambda}{2} \pm \sqrt{(E_{\Lambda,x} - E_{\Lambda,z} - \lambda_\Lambda)^2 + 2\lambda^2_\Lambda}}{2\sqrt{N_\pm}} \tag{3.28}$$

$$v^\Lambda_\pm = \frac{\lambda_\Lambda}{\sqrt{2N_\pm}} \tag{3.29}$$

with $N_\pm = \lambda_\Lambda^2 + \frac{1}{2}(E_x^\Lambda - E_z^\Lambda - \lambda_\Lambda)^2 \pm \frac{1}{2}(E_x^\Lambda - E_z^\Lambda - \lambda_\Lambda)\sqrt{(E_x^\Lambda - E_z^\Lambda - \lambda_\Lambda)^2 + 2\lambda_\Lambda^2}$. The energies are given by:

$$E_{\frac{3}{2}}^\Lambda = E_x^\Lambda + \frac{\lambda_\Lambda}{2} \tag{3.30}$$

$$E_{\frac{1}{2}}^{\Lambda\pm} = \frac{E_{\Lambda,x} + E_{\Lambda,z} - \frac{\lambda_\Lambda}{2} \pm \sqrt{(E_{\Lambda,x} - E_{\Lambda,z} - \lambda_\Lambda)^2 + 2\lambda_\Lambda^2}}{2} \tag{3.31}$$

The new energies are plotted in figure 3.6 as a function of the spin-orbit constant. We see that the $|\Lambda_{x,y}\rangle$ levels are split into 2 doubly degenerate levels (time reversal and inversion symmetry requires at least double degeneracy). The levels $|\Lambda_z\rangle$ degeneracy is pushed away from the other levels with the same Λ. Thus the two states closest to the chemical potential are pushed towards each other, and for a certain spin-orbit coupling strength, the two level crosses, creating a band inversion.

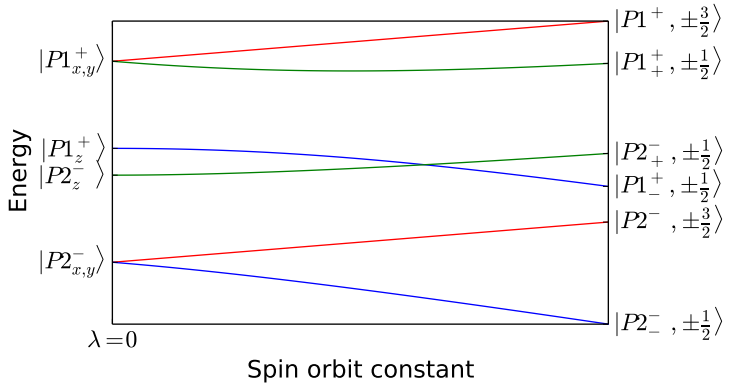


Figure 3.6: The energy levels as a function of the spin-orbit coupling constant. Here we have the same spin-orbit constant for the two bands $\lambda = \lambda_{P1^+} = \lambda_{P1^-}$. Note that for a strong enough spin-orbit coupling there is a level crossing allowing the material to go into a topologically non-trivial phase. The superscript sign on the basis states denotes the inversion eigenvalue. The spin $\frac{1}{2}$ states have an subscript, indicating which of the solutions to the 2×2 block $h_{\Lambda,\pm\frac{1}{2}}$ it is.

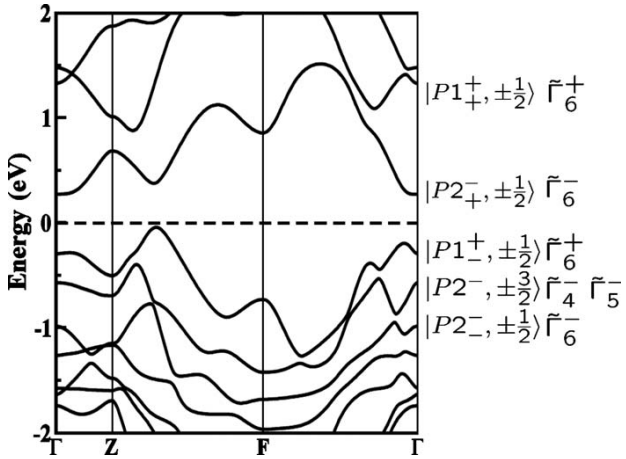


Figure 3.7: The band structure of bismuth selenide, obtained from ab initio calculations in [11]. The states are labeled

According to [11] the states $|P1^+_-, \pm\frac{1}{2}\rangle, |P2^-_+, \pm\frac{1}{2}\rangle$ turn out to be the states closest around the Fermi level at the gamma point, from ab initio calculations. See figure 3.7. Now we have seen how this fact qualitatively can be described by considering the nearest neighbour coupling, the crystal field splitting and the spin-orbit coupling. Specifically we saw that the spin-orbit coupling lead to a crossing of the two levels.

3.3 Model Hamiltonian

Now we have discussed qualitatively the effects of the couplings within one quintuple layer, the crystal field and the spin-orbit coupling. We saw how we ended with the two doubly degenerate energy levels $|P1^+_-, \pm\frac{1}{2}\rangle$ and $|P2^-_+, \pm\frac{1}{2}\rangle$ being closest to to the Fermi level. These states are eigenstates at the gammapoint. We will now construct an effective Hamiltonian, in the basis of these states. This is an example of *quasi degenerate perturbation theory* in $\mathbf{k}$. We will use the symmetries of the crystal to construct all allowed terms up to third order in $\mathbf{k}$.

In general any 4×4 hermitian matrix can be expanded in terms of the Dirac gamma matrices:

$$H_{eff}(\mathbf{k}) = \varepsilon(\mathbf{k}) + \sum_i d_i(\mathbf{k})\Gamma_i + \sum_{ij} d_{ij}(\mathbf{k})\Gamma_{ij}, \tag{3.32}$$

for real functions $d_i(\mathbf{k})$, $d_{ij}(\mathbf{k})$, for $i, j \in \{1, 2, 3, 4, 5\}$. Here we chose the Dirac matrices:

$$\Gamma_1 = \sigma_x \otimes \tau_x, \quad \Gamma_2 = \sigma_y \otimes \tau_x, \quad \Gamma_3 = \sigma_z \otimes \tau_x, \quad \Gamma_4 = \sigma_0 \otimes \tau_y, \quad \Gamma_5 = \sigma_0 \otimes \tau_z \tag{3.33}$$

and their commutators $\Gamma_{ij} = \frac{1}{2i}[\Gamma_i, \Gamma_j]$. We choose the basis $|P1^+_-, \frac{1}{2}\rangle$, $|P2^-_+, \frac{1}{2}\rangle$, $|P1^+_-, -\frac{1}{2}\rangle$, $|P2^-_+, -\frac{1}{2}\rangle$. The σ and τ matrices are two sets of pauli matrices in the space of total angular momentum and the $P1^+_-$, $P2^+_-$ subspaces, respectively.

First, we will need the matrices representing the symmetry operations of the crystal in this basis.

1. Inversion: All the basis states are eigenstates of the inversion operator labeled by their eigenvalue, thus $I = \sigma_0 \otimes \tau_z$.

2. Threefold rotation: Since three-fold rotation around the z axis does not change the positions of the atoms within one unit cell it can be written in terms of the z component of the angular momentum operator:

$$R_3 = e^{i\frac{2\pi}{3}J_z} = e^{i\frac{\pi}{3}\sigma_z \otimes \tau_0} = \begin{pmatrix} e^{i\pi/3} & 0 \\ 0 & e^{-i\pi/3} \end{pmatrix} \otimes \tau_0 \tag{3.34}$$

3. Twofold rotation: This rotation changes the Bi1 and Se1 sites into Bi1′ and Se1′ like inversion, so in the τ subspace, it is represented by τ_z. In the angular momentum subspace it is given by $e^{i\frac{\pi}{2}} = i\sigma_x$. Hence $R_2 = i\sigma_x \otimes \tau_z$.

4. Time reversal: This is not part of the group D_{3d} but should still be a symmetry of our final Hamiltonian. As we saw in section 2.9 for a spin $\frac{1}{2}$ system it can be represented by

$T = i\sigma_y K$, where K is complex conjugation. Therefore, we have $T = i\sigma_y \otimes \tau_0 K$.

We want an effective Hamiltonian which is invariant under all these transformations, i.e.

$$H_{eff}(\mathbf{k}) = U^{-1} H_{eff}(\mathbf{k}') U \tag{3.35}$$

where U is any of the transformation matrices, and $\mathbf{k}'$ is the transformed momentum, under the according symmetry transformation. We have chosen a coordinate system, such that k_z is perpendicular to the layers of the material while k_x and k_y are in the layers. The effect of the symmetry transformations on $\mathbf{k}$ are:

1. Inversion takes $\mathbf{k} \to -\mathbf{k}$
2. Threefold rotation around the z axis leaves k_z invariant but transforms:

$$\begin{pmatrix} k_x \\ k_y \end{pmatrix} \to \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} k_x \\ k_y \end{pmatrix} \tag{3.36}$$

   where $\theta = \frac{2\pi}{3}$
3. Twofold rotation around the x axis leaves k_x invariant, but transforms $k_y, k_z \to -k_y, -k_z$.
4. Time reversal takes $\mathbf{k} \to -\mathbf{k}$.

Therefore k_x, k_y transform according to the irreducible representation $\Gamma^{(3-)}$ of the point group D_{3d}, while k_z transforms according to $\Gamma^{(2-)}$. In addition k_x, k_y, k_z are odd under time reversal. From the transformation behaviour of k_x, k_y, k_z, we can find the transformation behaviour of higher order polynomials of k_x, k_y, k_z. This information is summarized in table 3.1.

Then we can work out how the Dirac matrices transform, by computing:

$$U\Gamma_a U^{-1}, \tag{3.37}$$

where a is either i or i, j, for $i, j \in \{1, 2, 3, 4, 5\}$. Then we expand the resulting matrices in the Dirac matrices again. The Dirac matrices were chosen such that they all transform according to some irreducible representation. If we take, for example, Γ_1 and Γ_2, and check their transformation properties we get:

1. Inversion: $\Gamma_i \to I\Gamma_i I^{-1} = -\Gamma_i$ for $i \in \{1, 2\}$.
2. Threefold rotation around the z axis:

$$\begin{pmatrix} -\Gamma_2 \\ \Gamma_1 \end{pmatrix} \to \begin{pmatrix} -R_3\Gamma_2 R_3^{-1} \\ R_3\Gamma_1 R_3^{-1} \end{pmatrix} = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix} \begin{pmatrix} -\Gamma_2 \\ \Gamma_1 \end{pmatrix} \tag{3.38}$$

3. Twofold rotation around the x axis: $\Gamma_1 \to R_2\Gamma_1 R_2^{-1} = -\Gamma_1$ and $\Gamma_2 \to R_2\Gamma_1 R_2^{-1} = \Gamma_2$.
4. Time reversal: $\Gamma_i \to T\Gamma_i T^{-1} = -\Gamma_i$ for $i \in \{1, 2\}$.

	$\Gamma^{(1+)}$	$\Gamma^{(2+)}$	$\Gamma^{(3+)}$	$\Gamma^{(1-)}$	$\Gamma^{(2-)}$	$\Gamma^{(3-)}$
T even	1 k_z^2 $k_x^2+k_y^2$ Γ_5		$\{2k_xk_y, k_x^2-k_y^2\}$ $\{k_yk_z, -k_xk_z\}$	Γ_{35}	Γ_{45}	$\{-\Gamma_{25}, \Gamma_{15}\}$
T odd		B_z Γ_{12} Γ_{34}	$\{B_x, B_y\}$ $\{\Gamma_{23}, \Gamma_{31}\}$ $\{\Gamma_{14}, \Gamma_{24}\}$	$k_x^3-3k_xk_y^2$ Γ_3	k_z k_z^3 $k_z(k_x^2+k_y^2)$ $k_y^3-3k_xk_y^2$ Γ_4	$\{k_x, k_y\}$ $\{k_xk_z^2, k_yk_z^2\}$ $\{k_x(k_x^2+k_y^2), k_y(k_x^2+k_y^2)\}$ $\{k_z(k_x^2-k_y^2), -2k_xk_yk_z\}$ $\{-\Gamma_2, \Gamma_1\}$

Table 3.1: All polynomials of momentum up to third order, categorized according to their reducible representation in the group D_{3d}. The Dirac matrices are listed the same way, such that terms in the same cell can be combined to form invariant terms in our model Hamiltonian. For the two dimensional representations we list pairs, transforming together with the notation $\{\cdot,\cdot\}$. The elements are listed such that they transform exactly the same way, and invariant terms can be formed by simply taking the dot product of two lists in the same cell. The components of a magnetic field is also listed, and using the symmetry principles, we can write down a Zeeman term.

Therefore, the set of matrices $\{-\Gamma_2, \Gamma_1\}$ transforms exactly like $\{k_x, k_y\}$. The transformation properties of all the Dirac matrices are summarized in table 3.1.

As we saw in section 2.11 invariant terms can only be formed by combinations of objects belonging to the same irreducible representation. Therefore, in the present case we can form an invariant term by combining $\{-\Gamma_2, \Gamma_1\}$ and $\{k_x, k_y\}$.

$$-\Gamma_2 k_x + \Gamma_1 k_y \tag{3.39}$$

By proceeding in this way, taking all combinations of Dirac matrices and polynomials transforming the same way (i.e. are listed in the same cell in table 3.1), we can find the most general Hamiltonian allowed by the symmetries of the crystal. Going to third order in $\mathbf{k}$ we arrive at the effective Hamiltonian:

$$\begin{aligned} H'_{eff} &= \varepsilon_0(\mathbf{k}) + \mathcal{M}(\mathbf{k})\Gamma_5 + \mathcal{A}_1(\mathbf{k})k_z\Gamma_4 + \mathcal{A}_2(\mathbf{k})(k_y\Gamma_1 - k_x\Gamma_2) \\ &\quad + R_1(k_x^3 - 3k_xk_y^2)\Gamma_3 + R_2(3k_x^2k_y - k_y^3)\Gamma_4 + R_3k_z(2k_xk_y\Gamma_1 + (k_x^2-k_y^2)\Gamma_2) \end{aligned} \tag{3.40}$$

where

$$\varepsilon_0(\mathbf{k}) = C + D_1k_z^2 + D_2k_{||}^2 \tag{3.41}$$

$$\mathcal{M}(\mathbf{k}) = M - B_1k_z^2 - B_2k_{||}^2 \tag{3.42}$$

$$\mathcal{A}_1(\mathbf{k}) = A_1 + A_{1,z}k_z^2 + A_{1,||}k_{||}^2 \tag{3.43}$$

$$\mathcal{A}_2(\mathbf{k}) = A_2 + A_{2,||}k_{||}^2 + A_{2,z}k_z^2 \tag{3.44}$$

where $k_{||} = \sqrt{k_x^2+k_y^2}$. The parameters $C, D_1, D_2, A_1, A_{1,||}, A_{1,z}, A_2, A_{2,||}, A_{2,z}, R_1, R_2, R_3$ are not given by symmetry considerations, and need to be determined either by fitting to ab initio calculations of the spectrum or by $k\cdot p$ theory using ab initio calculations of the wave

functions at the gamma point. We note here that the terms with coefficients $A_{1,||}$, $A_{2,z}$ and R_3 are not included in [11]. Whether they find these terms to be zero, or they are excluded for some other reason is not clear. The three terms have in common, that they are all third order terms, with a combination of the in plane momenta k_x, k_y and the out of plane momentum k_z.

This Hamiltonian looks different from the one used for example in [12] and [16], but is equivalent by a unitary transformation:

$$U_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & i \end{pmatrix} \tag{3.45}$$

under this unitary transformation we get the Hamiltonian:

$$H_{eff} = U_1 H'_{eff} U_1^\dagger = \varepsilon_0(\mathbf{k}) + \begin{pmatrix} \mathcal{M}(\mathbf{k}) & \mathcal{A}_1(\mathbf{k})k_z & 0 & \mathcal{A}_2(\mathbf{k})k_- \\ \mathcal{A}_1(\mathbf{k})k_z & -\mathcal{M}(\mathbf{k}) & \mathcal{A}_2(\mathbf{k})k_- & 0 \\ 0 & \mathcal{A}_2(\mathbf{k})k_+ & \mathcal{M}(\mathbf{k}) & -\mathcal{A}_1(\mathbf{k})k_z \\ \mathcal{A}_2(\mathbf{k})k_+ & 0 & -\mathcal{A}_1(\mathbf{k})k_z & -\mathcal{M}(\mathbf{k}) \end{pmatrix}$$
$$+\frac{R_1}{2}\begin{pmatrix} 0 & i(R_d k_+^3 + R_m k_-^3) & 0 & -R_3 k_z k_+^2 \\ -i(R_d k_-^3 + R_m k_+^3) & 0 & -R_3 k_z k_+^2 & 0 \\ 0 & -R_3 k_z k_-^2 & 0 & i(R_m k_+^3 + R_d k_-^3) \\ -R_3 k_z k_-^2 & 0 & -i(R_m k_+^3 + R_d k_-^3) & 0 \end{pmatrix} \tag{3.46}$$

where we have introduced two new parameters $R_m = \frac{R_1+R_2}{2}$ and $R_d = \frac{R_1-R_2}{2}$ and $k_\pm = k_x \pm k_y$ to make the notation less messy. The second term includes third order terms that are just first order terms multiplied by invariant second order terms, while the third term contains the rest of the allowed third order terms. It is important to note, as mentioned in [17], that this unitary transformation is affecting the spin operators:

$$s_x = \frac{1}{2} U_1 \sigma_x \otimes \tau_0 U_1^\dagger = \frac{1}{2}\sigma_x \otimes \tau_z \tag{3.47}$$

$$s_y = \frac{1}{2} U_1 \sigma_y \otimes \tau_0 U_1^\dagger = \frac{1}{2}\sigma_y \otimes \tau_z \tag{3.48}$$

$$s_z = \frac{1}{2} U_1 \sigma_z \otimes \tau_0 U_1^\dagger = \frac{1}{2}\sigma_z \otimes \tau_0 \tag{3.49}$$

This will of course be very important when we calculate the spin structure of the surface states. We emphasize here that when we talk about spin in this model, it is actually the total angular momentum of the electronic states.

3.3.1 Magnetic field

If a magnetic field is present it affects the Hamiltonian in two ways; the orbital effect and a Zeeman term. The orbital effect can be included by Peierls substitution $k \to k + e\mathbf{A}$, where $\mathbf{A}$ is the vector potential. The form of the Zeeman term can be deduced by requiring invariance

C(eV)	-0.0083	-0.0068
D_1(eVÅ^2)	5.74	1.3
D_2(eVÅ^2)	30.4	19.6
M(eV)	-0.28	0.28
B_1(eVÅ^2)	-6.86	10.0
B_2(eVÅ^2)	-44.5	56.6
A_1(eV Å)	2.26	2.2
A_2(eV Å)	3.33	4.1
R_1(eVÅ^3)	50.6	
R_2(eVÅ^3)	-113.3	
g_{1z}	-25.4	
g_{1p}	-4.12	
g_{2z}	4.10	
g_{2p}	4.80	

Table 3.2: Parameters for the effective model. The first column is calculated from $k \cdot p$ theory in [11] using ab initio calculations of wavefunctions, while the second is from [12] from fitting the spectrum to that of an ab initio calculation.

under D_{3d} and time reversal. Here we use the convention that the magnetic field is reversed by time reversal operation, therefore the Zeeman term should be invariant under time reversal. Using table 3.1 as before, we get:

$$\begin{aligned} H'_z &= (\alpha_1\Gamma_{12} + \alpha_2\Gamma_{34})B_z + \alpha_3(B_x\Gamma_{23} + B_y\Gamma_{31}) + \alpha_4(B_x\Gamma_{14} + B_y\Gamma_{24}) \\ &= \begin{pmatrix} (\alpha_1+\alpha_2)B_z & 0 & (\alpha_3+\alpha_4)B_- & 0 \\ 0 & (\alpha_1-\alpha_2)B_z & 0 & (\alpha_3-\alpha_4)B_- \\ (\alpha_3+\alpha_4)B_+ & 0 & -(\alpha_1+\alpha_2)B_z & 0 \\ 0 & (\alpha_3-\alpha_4)B_+ & 0 & -(\alpha_1-\alpha_2)B_z \end{pmatrix} \\ &= \frac{\mu_B}{2}\begin{pmatrix} g_{1z}B_z & 0 & g_{1p}B_- & 0 \\ 0 & g_{2z}B_z & 0 & g_{2p}B_- \\ g_{1p}B_+ & 0 & -g_{1z}B_z & 0 \\ 0 & g_{2p}B_+ & 0 & -g_{2z}B_z \end{pmatrix} \end{aligned} \tag{3.50}$$

The new parameters $g_{1z} = \frac{2}{\mu_B}(\alpha_1 + \alpha_2)$, $g_{2z} = \frac{2}{\mu_B}(\alpha_1 - \alpha_2)$, $g_{1p} = \frac{2}{\mu_B}(\alpha_3 + \alpha_4)$, $g_{2p} = \frac{2}{\mu_B}(\alpha_3 - \alpha_4)$ are effective g factors in the different orbitals. If we again perform the unitary transformation we get the Zeeman term:

$$H_z = U_1 H'_z U_1^\dagger = \frac{\mu_B}{2}\begin{pmatrix} g_{1z}B_z & 0 & g_{1p}B_- & 0 \\ 0 & g_{2z}B_z & 0 & -g_{2p}B_- \\ g_{1p}B_+ & 0 & -g_{1z}B_z & 0 \\ 0 & -g_{2p}B_+ & 0 & -g_{2z}B_z \end{pmatrix} \tag{3.51}$$

3.4 Bulk states

Now we have arrived at the effective model in eq. (3.46), which was the main goal of this chapter. This Hamiltonian will be the starting point for our investigation of the surface states

in bismuth selenide. But before we introduce a surface, it is instructive to find the bulk states of this Hamiltonian.

In the case of an infinite insulator, we can simply diagonalize this 4×4 Hamiltonian. We will do this only to second order in the momentum. Using the σ and τ matrices, our Hamiltonian can be written:

$$H(\mathbf{k}) = \varepsilon_0(\mathbf{k})\mathcal{M}(\mathbf{k})\sigma_0 \otimes \tau_z + A_2(k_x\sigma_x + k_y\sigma_y) \otimes \tau_x + A_1 k_z \sigma_z \otimes \tau_x. \tag{3.52}$$

The spectrum can be found easily by squaring $H(\mathbf{k}) - \varepsilon_0(\mathbf{k})$ and using the anticommutativity of the paulimatrices which cancels all the crossterms, i.e.

$$\begin{aligned}(H(\mathbf{k}) - \varepsilon_0(\mathbf{k}))^2 &= (\mathcal{M}(\mathbf{k})\sigma_0 \otimes \tau_z + A_2(k_x\sigma_x + k_y\sigma_y) \otimes \tau_x + A_1 k_z \sigma_z \otimes \tau_x)^2 \\ &= (\mathcal{M}(\mathbf{k})^2 + A_1^2 k_{||}^2 + A_2^2 k_z^2).\end{aligned} \tag{3.53}$$

Hence, $(H(\mathbf{k}) - \varepsilon_0(\mathbf{k}))^2$ is proportional to the identity matrix and has only one eigenvalue $(\mathcal{M}(\mathbf{k})^2 + A_1^2 k_{||}^2 + A_2^2 k_z^2)$. The eigenvalues of a squared matrix are simply the squares of the eigenvalues of the original matrix. Therefore, the matrix $H(\mathbf{k}) - \varepsilon_0(\mathbf{k})$ must have eigenvalues $\pm\sqrt{\mathcal{M}(\mathbf{k})^2 + A_1^2 k_{||}^2 + A_2^2 k_z^2}$. Then we get the bulk spectrum:

$$E^{\pm}_{\text{bulk}} = \varepsilon_0(\mathbf{k}) \pm \sqrt{\mathcal{M}(\mathbf{k})^2 + A_1^2 k_z^2 + A_2^2 k_{||}^2}. \tag{3.54}$$

We get two doubly degenerate bands due to the combination of time reversal and inversion symmetry, as described in section 2.9.2. The dispersion relation is plotted in figure 3.8. The momentum dependent terms in $\varepsilon_0(\mathbf{k})$ breaks the particle-hole symmetry. If these are not included the conduction band and valence band would be exactly symmetric around the value of C.

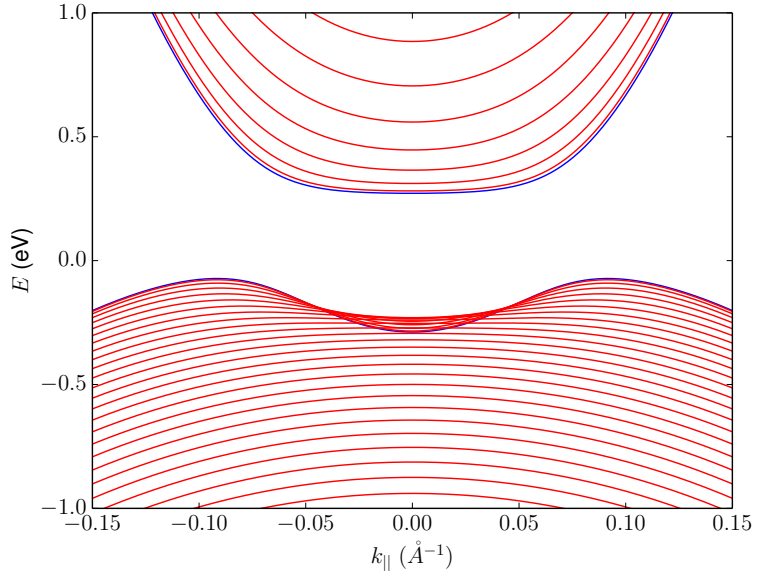


Figure 3.8: The energy of the eigenstates in the low energy model, as a function of the in plane momentum $k_{||} = \sqrt{k_x^2 + k_y^2}$ for evenly spaced values of k_z (with a spacing of $0.03\,\text{Å}^{-1}$). The blue one is for $k_z = 0$. Here we used the parameters from [11].

The eigenstates are found to be:

$$\psi_{1\pm}(\mathbf{k}) = C\begin{pmatrix}\mathcal{M}(\mathbf{k}) \mp \sqrt{\mathcal{M}(\mathbf{k})^2 + A_1^2 k_{||}^2 + A_2^2 k_z^2}\\ A_1 k_z \\ 0 \\ A_2 k_+\end{pmatrix}, \psi_{2\pm}(\mathbf{k}) = C\begin{pmatrix}0\\ A_2 k_- \\ \mathcal{M}(\mathbf{k}) \mp \sqrt{\mathcal{M}(\mathbf{k})^2 + A_1^2 k_{||}^2 + A_2^2 k_z^2} \\ -A_1 k_z\end{pmatrix}. \tag{3.55}$$

where C is a normalization constant. The upper/lower sign corresponds to the upper/lower band. We see that these states are Kramers doublets $T\psi_{1\pm}(\mathbf{k}) = -\psi_{2\pm}(-\mathbf{k})$. The eigenstates can be chosen in many different ways because of the degeneracy, but the choice is not neccesarrily valid for any $\mathbf{k}$, since the eigenvectors could become linearly dependent at som special momentum. The four eigenvectors are linearly independent if the determinant of the matrix $(\psi_{1+}, \psi_{1-}, \psi_{2+}, \psi_{2-}) \neq 0$. The determinant of the vectors chosen here is:

$$\det(\psi_{1+}, \psi_{1-}, \psi_{2+}, \psi_{2-}) = 4A_2^4 k_{||}^2 + 8A_1^2 A_2^2 k_{||}^2 k_z^2 + 4A_1^4 k_z^4 + 4A_2^2 k_{||}^2 \mathcal{M}(\mathbf{k})^2 + 4A_1^2 k_z^2 \mathcal{M}(\mathbf{k})^2. \tag{3.56}$$

This is seen to be zero only at $\mathbf{k} = 0$ (if we take A_1 and A_2 to be nonzero), and at that point the Hamiltonian is diagonal and is trivially solved.

3.5 Envelope function approximation

The full wave functions of the bulk states found here, are given by an expansion in the basis states of our model. In general an eigenspinor of our model at momementum $\mathbf{k}$, denoted by:

$$\psi(\mathbf{k}) = \begin{pmatrix}a\\ b\\ c\\ d\end{pmatrix}, \tag{3.57}$$

corresponds to the wave function:

$$\Psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}\left(a\langle\mathbf{r}|P1^+_-, \tfrac{1}{2}\rangle + b\langle\mathbf{r}|P2^-_+, \tfrac{1}{2}\rangle + c\langle\mathbf{r}|P1^+_-, -\tfrac{1}{2}\rangle + d\langle\mathbf{r}|P2^-_+, -\tfrac{1}{2}\rangle\right). \tag{3.58}$$

The basis state wave function, contains the internal structure, within one quintuple layer, whereas the plane wave factor $e^{i\mathbf{k}\cdot\mathbf{r}}$ acts as an envelope function. In the following chapters, when we introduce one or two surfaces, the translational symmetry in the z direction is broken. Therefore, k_z is no longer a good quantum number, and we will make the substitution $k_z \rightarrow$

$-i\partial_z$. In general a solution to our model becomes z dependent

$$\psi_{k_x,k_y}(z) = \begin{pmatrix} a(z) \\ b(z) \\ c(z) \\ d(z) \end{pmatrix}, \tag{3.59}$$

corresponding to the full wave function:

$$\Psi(\mathbf{r}) = e^{ik_x x + ik_y y} \left(a(z)\langle \mathbf{r}|P1^+_-, \tfrac{1}{2}\rangle + b(z)\langle \mathbf{r}|P2^-_+, \tfrac{1}{2}\rangle + c(z)\langle \mathbf{r}|P1^+_-, -\tfrac{1}{2}\rangle + d(z)\langle \mathbf{r}|P2^-_+, -\tfrac{1}{2}\rangle \right). \tag{3.60}$$

We still have the lattice periodic basis functions, and plane wave envelope functions in the x and y directions. In the z direction we have an envelope function, which is determined by the boundary conditions. This is a way to seperate length scales, the lattice periodic functions, gives the structure on an atomic scale, whereas the envelope function gives the structure on larger scales due to the boundaries. For a more detailed discussion on the envelope function approximation, see [18]

In the chapters 4 and 5, we will work entirely within this approximation, simply refering to the envelope functions as *the* wave functions.

Surface states on a single surface Bi_2Se_3

In this chapter we will discuss a so-called semi-infinite geometry, which is an infinite insulator in the x and y directions with a surface perpendicular to the z-axis at $z = 0$. We will take the the topological insulator to be filling the $z < 0$ half-space, see figure 4.1. We will use the model Hamiltonian 3.46 to second order in the momenta. Even though a finite geometry with two surfaces is more realistic and interesting, the semi-infinite case can be solved analytically and thus gives a nice clear picture. We will find the criteria for the existence of surface states based on the parameters of the Hamiltonian. Furthermore we will characterize the spatial and spin structure of the found surface states.

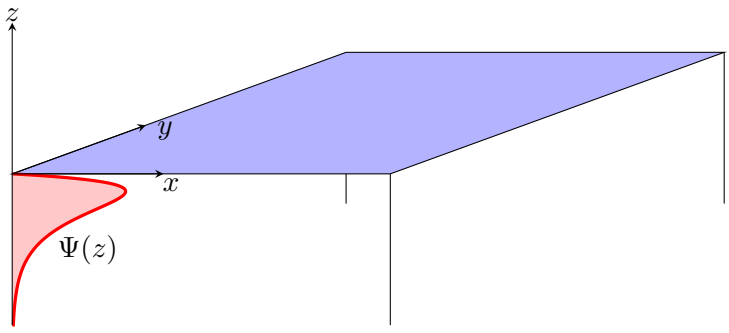


Figure 4.1: The system considered in this chapter is an semi-infinite topological insulator filling the $z < 0$ half-space. This gives rise to localized surface states close to the surface at $z = 0$.

To find the surface states we need to impose some boundary conditions. Our choice is the so-called hard wall or open boundary conditions:

$$\Psi(z = 0) = 0 \tag{4.1}$$

To get localized states, we also need the wave function to decay away from the surface, i.e.:

$$\Psi(z \to -\infty) = 0 \tag{4.2}$$

Other boundary conditions could be imposed, in [19] a wider class of boundary conditions, including open boundary conditions, are explored. They find that both the spectrum and the existence of surface states are dependent on the boundary condition, as well as the parameters of the model. Another method is to let the model be valid in all space, but let the parameter M change sign across the surface. These considerations however, are beyond the scope of this thesis, where we will only use the open boundary conditions.

4.1 Ansatz

If we have a surface perpendicular to the z axis, the translational symmetry in this direction is broken, and k_z is no longer a good quantum number. We will therefore use a more general approach, where we restore the differential operators in the Hamiltonian $k_z \to -i\partial_z$ giving:

$$H(k_x, k_y, -i\partial_z) = \varepsilon_0(k_{||}, -i\partial_z) \tag{4.3}$$

$$+ \begin{pmatrix} \mathcal{M}(k_{||}, -i\partial_z) & -iA_1\partial_z & 0 & A_2k_- \\ -iA_1\partial_z & -\mathcal{M}(k_{||}, -i\partial_z) & A_2k_- & 0 \\ 0 & A_2k_+ & \mathcal{M}(k_{||}, -i\partial_z) & iA_1\partial_z \\ A_2k_+ & 0 & iA_1\partial_z & -\mathcal{M}(k_{||}, -i\partial_z) \end{pmatrix}, \tag{4.4}$$

$$\varepsilon_0(k_{||}, -i\partial_z) = C - D_1\partial_z^2 + D_2k_{||}^2, \tag{4.5}$$

$$\mathcal{M}(k_{||}, -i\partial_z) = M + B_1\partial_z^2 - B_2k_{||}^2, \tag{4.6}$$

$$k_\pm = k_x \pm ik_y, \tag{4.7}$$

$$k_{||} = \sqrt{k_x^2 + k_y^2}. \tag{4.8}$$

The time independent Schrödinger equation;

$$H(k_x, k_y, -i\partial_z)\Phi_{k_x,k_y}(x, y, z) = E\Phi_{k_x,k_y}(x, y, z), \tag{4.9}$$

where

$$\Phi_{k_x,k_y}(x, y, z) = \frac{1}{\sqrt{L_xL_y}} e^{ik_xx+ik_yy}\Psi_{k_x,k_y}(z) \tag{4.10}$$

since k_x and k_y are still good quantum numbers. The z dependent part $\Psi(z)$ is some, yet unknown 4-spinor. Since the x and y dependence is completely trivial, we will leave it out for most of this chapter, and simply treat this as a one dimensional problem, given by:

$$H(k_x, k_y, -i\partial_z)\Psi_{k_x,k_y}(z) = E\Psi_{k_x,k_y}(z). \tag{4.11}$$

This is now a system of four coupled second order differential equations in one variable, thus the space of solutions is 8 dimensional. Since it is a linear homogeneous system with constant coefficients, we can use the ansatz $\Psi(z) = \psi_\lambda e^{\lambda z}$, where ψ_λ is some z independent 4-vector. If we by this ansatz can find 8 independent solutions we know that they span the entire space of solutions. Hence, a general eigenfunction to 4.4, can be written as a linear combination of these 8 independent solutions. Then, by imposing the relevant boundary conditions, one can find the coefficients of this linear combination.

4.1.1 Eigenstates

Now we will find the eigenstates in the ansatz $\Psi(z) = \psi_\lambda e^{\lambda z}$, following the derivation in [16]. When the hamiltionian works on this ansatz, we get the replacement $\partial_z \to \lambda$. The Schrödinger

equation;

$$H(k_x, k_y, -i\lambda)\psi_\lambda = E\psi_\lambda \tag{4.12}$$

has non-trivial solutions when:

$$\begin{aligned}
0 &= \det(H(k_x, k_y, -i\lambda) - E) && (4.13)\\
\Leftrightarrow 0 &= \left(\mathcal{M}(k_{||}, -i\lambda)^2 - (\varepsilon_0(k_{||}, -i\lambda) - E)^2 + A_2^2 k_{||}^2 - A_1^2\lambda^2\right)^2 \\
\Leftrightarrow 0 &= \mathcal{M}(k_{||}, -i\lambda)^2 - (\varepsilon_0(k_{||}, -i\lambda) - E)^2 + A_2^2 k_{||}^2 - A_1^2\lambda^2 \\
\Leftrightarrow 0 &= (M + B_1\lambda^2 - B_2 k_{||}^2)^2 - (C - D_1\lambda^2 + D_2 k_{||}^2 - E)^2 + A_2^2 k_{||}^2 - A_1^2\lambda^2 \\
\Leftrightarrow 0 &= (B_1^2 - D_1^2)\lambda^4 + \left[2B_1(M - B_2 k_{||}^2) - 2D_1(C + D_2 k_{||}^2 - E) - A_1^2\right]\lambda^2 \\
&\quad + (M - B_2 k_{||}^2)^2 - (C + D_2 k_{||}^2 - E)^2 + A_2^2 k_{||}^2 \\
\Leftrightarrow 0 &= D_+ D_- \lambda^4 + F\lambda^2 + (E - L_1)(E - L_2) - A_2^2 k_{||}^2 && (4.14)\\
\Leftrightarrow \lambda &= \beta\lambda_\alpha = \beta\sqrt{\frac{-F + (-1)^{\alpha-1}\sqrt{R}}{2D_+ D_-}} && (4.15)
\end{aligned}$$

where $\beta = \pm$ and $\alpha \in \{1, 2\}$ and we have defined:

$$\begin{aligned}
F &= A_1^2 + D_+(E - L_1) + D_-(E - L_2), && (4.16)\\
R &= F^2 - 4D_+ D_- \left((E - L_1)(E - L_2) - A_2^2 k_{||}^2\right), && (4.17)\\
D_\pm &= D_1 \pm B_1, && (4.18)\\
L_1 &= C + M + (D_2 - B_2)k_{||}^2, && (4.19)\\
L_2 &= C - M + (D_2 + B_2)k_{||}^2. && (4.20)
\end{aligned}$$

The square root here denotes the principal value, to make λ_α uniquely defined. [1]

This gives four different solutions for λ for a given energy E. We could also have found the energy from the secular equation, equation 4.13, giving:

$$E = \varepsilon_0(k_{||}, \lambda) \pm \sqrt{A_2^2 k_{||}^2 - A_1^2\lambda^2 + \mathcal{M}(k_{||}, -i\lambda)^2},$$

which is simply the bulk spectrum with the replacement $k_z \to -i\lambda$. Usually one uses the secular equation to find the energy, but the real question here is; given some $k_{||}$ and E is it possible to make a superposition of $\psi_\lambda e^{\lambda z}$ that fulfills the boundary condition. If we had an infinite insulator the only condition on the wave function is that it does not diverge, which means that λ must be purely imaginary, thus we can simply plug any imaginary λ into the equation for the energy, and get the bulk spectrum again. Note that here we have assumed that $D_+ D_- \neq 0$, which will be justified later.

Equation (4.15) is extremely important and is the equation that is going to tell us whether or not surface states are possible. It tells us given some energy E and $k_{||}$ how the wave function can vary in the z direction. If there is only purely imaginary solutions for λ the wave function

[1] The principal value of the square root is the one with positive real part, and if the real part is zero the one with positive imaginary part.

will just oscillate, and if there is a real part it might be possible to find a solution that decays away from a surface. Whether it is possible will depend on the boundary conditions and how the actual eigenvectors ψ_λ turn out. Since the spectrum is doubly degenerate for a given λ and corresponding energy E there is two independent eigenvectors, given by:

$$\psi_{\alpha\beta 1} = \begin{pmatrix} \mathcal{M}(k_{||}, -i\beta\lambda_\alpha) \pm \sqrt{A_2 k^2 - A_1\lambda_\alpha^2 + \mathcal{M}(k_{||}, -i\beta\lambda_\alpha)^2} \\ -iA_1\beta\lambda_\alpha \\ 0 \\ A_2 k_+ \end{pmatrix} = \begin{pmatrix} D_+\lambda_\alpha^2 - L_2 + E \\ -iA_1\beta\lambda_\alpha \\ 0 \\ A_2 k_+ \end{pmatrix}. \tag{4.21}$$

$$\psi'_{\alpha\beta 2} = \begin{pmatrix} -iA_1\beta\lambda_\alpha \\ -\mathcal{M}(k_{||}, -i\beta\lambda_\alpha) \pm \sqrt{A_2 k^2 - A_1\lambda_\alpha^2 + \mathcal{M}(k_{||}, -i\beta\lambda_\alpha)^2} \\ A_2 k_+ \\ 0 \end{pmatrix} = \begin{pmatrix} -iA_1\beta\lambda_\alpha \\ D_-\lambda_\alpha^2 - L_1 + E \\ A_2 k_+ \\ 0 \end{pmatrix}. \tag{4.22}$$

These two are independent unless $A_2 k_+ = 0$, but we also want to calculate the surface state at $k_{||} = 0$, so we will instead use $\psi_{\alpha\beta 1}$ and:

$$\psi_{\alpha\beta 2} = \frac{D_-\lambda_\alpha^2 - L_1 + E}{A_2 k_+}\psi_{\alpha\beta 1} + \frac{iA_1\beta\lambda_\alpha}{A_2 k_+}\psi'_{\alpha\beta 2} = \begin{pmatrix} A_2 k_- \\ 0 \\ iA_1\beta\lambda_\alpha \\ D_-\lambda_\alpha^2 - L_1 + E \end{pmatrix}, \tag{4.23}$$

which is now linearly independent from $\psi_{\alpha\beta 1}$ unless $A_1\lambda_\alpha = 0$. Therefore we will assume $A_1 \neq 0$. The first component in eq. (4.23) was simplified by using eq. (4.14) to calculate the product:

$$\begin{aligned} &(D_+\lambda^2 - L_2 + E)(D_+\lambda^2 - L_2 + E) && (4.24)\\ &= D_+D_-\lambda^4 + (D_+(E - L1) + D_-(E - L_2))\lambda^2 + (E - L_1)(E - L_2) && (4.25)\\ &= D_+D_-\lambda^4 + (F - A_1^2)\lambda^2 + (E - L_1)(E - L_2) && (4.26)\\ &= A_2^2 k_{||}^2 - A_1^2\lambda^2 && (4.27) \end{aligned}$$

Now a general solution to the Schrödinger equation can be written:

$$\Psi(E, k_{||}, z) = \sum_{\alpha,\beta,\gamma} C_{\alpha\beta\gamma}\psi_{\alpha\beta\gamma}e^{\beta\lambda_\alpha z}, \tag{4.28}$$

where $\beta \in \{+, -\}$ and $\alpha, \gamma \in \{1, 2\}$. Both the spinors, $\psi_{\alpha\beta\gamma}$, and λ_α depend on both E and $k_{||}$. The coefficients $C_{\alpha\beta\gamma}$ is to be determined from boundary conditions.

$F^2 - R$	R	$\frac{-F}{2D_+D_-}$	λ	Possible states
+	-	$\pm$	$\lambda_1, \lambda_2 \in \mathbb{C}$ and $\lambda_1 = \lambda_2^*$	Surface
+	+	+	$\lambda_1, \lambda_2 \in \mathbb{R}$	Surface
+	+	-	$\lambda_1, \lambda_2 \in \mathbb{I}$	Bulk
-	+	$\pm$	$\lambda_i \in \mathbb{R}, \lambda_j \in \mathbb{I}$	Bulk

Table 4.1: Classifications of the different possibilities for solutions for λ. R and F are functions of $k_{||}$ and E so this divides the $k_{||}, E$ plane into different regions where either surface or bulk states are possible. Note that this is includes all possible combinations, since $R < 0 \Rightarrow F^2 - R > 0$.

4.1.2 Spatial structure of the eigenstates

Before imposing the boundary conditions, and actually finding the surface states we will look at the spatial dependence of the found eigenstates.

The spatial structure of these solutions is determined by the $\lambda_\alpha(k_{||}, E)$. Surface states are only possible if both λ_1 and λ_2 have a real part, thus by analyzing the dependence of λ_α on $k_{||}$ and E it is possible to find out, where in the $(k_{||}, E)$ plane surface states are possible. The essential part is the sign of the functions R, $F^2 - R$ and $-\frac{F}{2D_+D_-}$. We realize this by looking at:

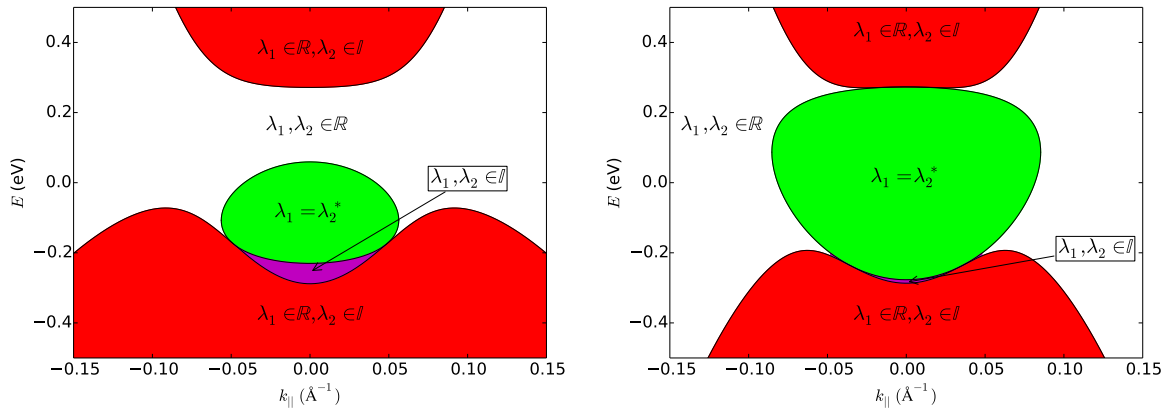
$$\lambda_\alpha^2 = \frac{-F + (-1)^\alpha \sqrt{R}}{2D_+D_-} \tag{4.29}$$

If $R < 0$, then $\sqrt{R}$ is imaginary then λ_α^2 is complex and so is λ_α. In addition, we see that λ_1 and λ_2 are complex conjugate partners. If $R > 0$, then λ_α^2 is real, and λ_α is either purely real or purely imaginary, depending on the sign of λ_α^2. If $F^2 - R > 0 \Rightarrow |F| > \sqrt{R}$ then λ_1^2 and λ_2^2 have the same sign, and λ_1 and λ_2 are both either real or imaginary, depending on the sign of $-\frac{F}{2D_+D_-}$. On the other hand if $F^2 - R < 0 \Rightarrow |F| < \sqrt{R}$ λ_1^2 and λ_2^2 have different signs, and we have on imaginary and one real λ_α, for $\alpha \in \{1, 2\}$. The different possibilities are summarized in table 4.1.

In figure 4.2 the $(k_{||}, E)$ plane is divided into regions of these four qualitative possibilities for λ_1 and λ_2. We recognize that the boundary of the region, where one of the λ_α's is real and the other imaginary, is exactly the $k_z = 0$ bulk spectrum. This must be true, since across this boundary one of the λ_α's change from purely imaginary to purely real, and since it is a continuous function it must be 0 on the boundary. Eigenstates with $\lambda = 0$ are of course equal to the bulk eigenstates with $k_z = 0$. The regions, where at least one λ_α, is imaginary there exists bulk states with $k_z = i\lambda_\alpha$. On the boundary of the region where both λ_α's are complex they are equal, because R goes continuously from negative to positive, and therefore $R = 0 \Leftrightarrow \lambda_1 = \lambda_2$ at the boundary[2]. At this boundary our ansatz actually breaks down, since we only have 4 independent solutions to our 4 second order differential equations.

If surface states exist for the situation in 4.2a, the spatial structure of the surface states depend on where in the $(k_{||}, E)$ plane they are located. If they are located in the green region, they will have some oscillatory behaviour because of the imaginary part of λ_1 and λ_2. In the

[2] Actually, when they go into the region where they are both imaginary there is some ambiguity because of the branch cut of the complex square root. But since the possible exponents in any case are $\pm\lambda_\alpha$ it is only a question of which 2 of the four solutions, we denote by λ_α.



(a) Parameters from [11].

(b) Parameters from [12]

Figure 4.2: The different regions in the $k_{||}, E$ plane. In the white and green regions only surface states are possible, while only bulk states are possible in the red and purple regions. The red region is actually the union of regions with $\lambda_1 \in \mathbb{R}, \lambda_2 \in \mathbb{I}$ and $\lambda_1 \in \mathbb{I}, \lambda_2 \in \mathbb{R}$, but since this has no physical significance we have just joined them.

white region both λ_1 and λ_2 are real, and a surface state in this region will have a simple decay. For the situation in figure 4.2b, we see that for $k_{||} = 0$ both λ_1 and λ_2 are complex, and the surface state will have oscillatory behaviour.

4.2 Existence and spectrum of surface states

After this qualitative discussion, on the possibilty of surface states, we will now find the find the quantitative criteria, for the existence of surface states, as well as the surface state spectrum. To do this we need to impose the boundary conditions:

$$\Psi_{k_x,k_y}(z=0) = 0 \text{ and } \Psi_{k_x,k_y}(z \to -\infty) = 0. \tag{4.30}$$

The second one tells us that the λ's must have a positive real part, and thus we can immediately drop the $\beta = -$ terms in the general solution in eq. (4.28). The condition at $z = 0$ gives the following equation for the remaining four coefficients:

$$0 = \Psi_{k_x,k_y}(z=0) = \sum_{\alpha,\gamma} C_{\alpha+\gamma}\psi_{\alpha+\gamma} = \begin{pmatrix}\psi_{1+1} & \psi_{1+2} & \psi_{2+1} & \psi_{2+2}\end{pmatrix} \begin{pmatrix} C_{1+1} \\ C_{1+2} \\ C_{2+1} \\ C_{2+2} \end{pmatrix} \tag{4.31}$$

$$\Leftrightarrow 0 = \begin{pmatrix} J_{1+} & A_2 k_- & J_{2+} & A_2 k_- \\ -iA_1\lambda_1 & 0 & -iA_1\lambda_2 & 0 \\ 0 & iA_1\lambda_1 & 0 & iA_1\lambda_2 \\ A_2 k_+ & J_{1-} & A_2 k_+ & J_{2-} \end{pmatrix} \begin{pmatrix} C_{1+1} \\ C_{1+2} \\ C_{2+1} \\ C_{2+2} \end{pmatrix} \tag{4.32}$$

where we defined $J_{\alpha+} = D_+\lambda_\alpha^2 - L_2 + E$ and $J_{\alpha-} = D_-\lambda_\alpha^2 - L_1 + E$. The secular equation for nontrivial solutions of the above system gives:

$$\begin{aligned} 0 &= iA_1\lambda_1\left(iA_1\lambda_2 J_{2+}J_{1-} + iA_1A_2^2\lambda_1 k_{||}^2 - iA_1A_2^2\lambda_2 k_{||}^2 - iA_1\lambda_1 J_{2-}J_{2+}\right) && (4.33)\\ &\quad + iA_1\lambda_2\left(iA_1\lambda_1 J_{1+}J_{2-} + iA_1A_2^2\lambda_2 k_{||}^2 - iA_1A_2^2\lambda_1 k_{||}^2 - iA_1\lambda_2 J_{1-}J_{1+}\right)\\ &= A_1^2\lambda_1\lambda_2\left(2A_2^2k_{||}^2 - J_{2+}J_{1-} - J_{1+}J_{2-}\right) && (4.34)\\ &\quad + A_1^2\lambda_1^2(J_{2-}J_{2+} - A_2^2k_{||}^2)\\ &\quad + A_1^2\lambda_2^2(J_{1-}J_{1+} - A_2^2k_{||}^2). \end{aligned}$$

This gives zero, if $A_1 = 0$. But for $A_1 = 0$ the basis vectors are no longer independent, and our method breaks down. Therefore, we assumed $A_1 \neq 0$, but this assumption will be justified later. To simplify eq. (4.34), we consider the product:

$$\begin{aligned} D_+D_-(\lambda_1^2 - \lambda_2^2)^2 &= (J_{1+} - J_{2+})(J_{1-} - J_{2-}) && (4.35)\\ &= J_{1+}J_{1-} + J_{2+}J_{2-} - J_{2+}J_{1-} - J_{1+}J_{2-} && (4.36)\\ &= 2A_2^2k_{||}^2 - A_1^2(\lambda_1^2 + \lambda_2^2) - J_{2+}J_{1-} - J_{1+}J_{2-} && (4.37)\\ \Leftrightarrow 2A_2^2k_{||}^2 - J_{2+}J_{1-} - J_{1+}J_{2-} &= D_+D_-(\lambda_1^2 - \lambda_2^2)^2 + A_1^2(\lambda_1^2 + \lambda_2^2), && (4.38) \end{aligned}$$

where we have used that $J_{\alpha+}J_{\alpha-} = A_2^2k_{||}^2 - A_1^2\lambda_\alpha^2$ from eq. (4.27). Inserting this in eq. (4.34) gives:

$$\begin{aligned} 0 &= A_1^2\lambda_1\lambda_2 D_+D_-(\lambda_1^2 - \lambda_2^2)^2 + A_1^4\lambda_1\lambda_2(\lambda_1^2 + \lambda_2^2) - 2A_1^4\lambda_1^2\lambda_2^2 && (4.39)\\ \Leftrightarrow 0 &= \frac{D_+D_-}{A_1^2}(\lambda_1^2 - \lambda_2^2)^2 + (\lambda_1^2 + \lambda_2^2) - 2\lambda_1\lambda_2\\ \Leftrightarrow 0 &= \frac{D_+D_-}{A_1^2}(\lambda_1 - \lambda_2)^2(\lambda_1 + \lambda_2)^2 + (\lambda_1 - \lambda_2)^2\\ \Leftrightarrow (\lambda_1 + \lambda_2)^2 &= -\frac{A_1^2}{D_+D_-} && (4.40) \end{aligned}$$

Remember that the λ's are functions of $k_{||}$ and E. If there exists a solution $E(k_{||})$ to eq. (4.40), then a wave function exists, using only the positive ($\beta = +$) square root solutions for λ, that fulfills the boundary condition at $z = 0$. Hence, a surface state exists with energy E and in plane momentum $k_{||}$. The square of a complex number $(a+ib)^2 = a^2 - b^2 + i2ab$ is real, only if the number is either purely real or imaginary. Since the right hand side in eq. (4.40) is real, it tells us that $\lambda_1 + \lambda_2$ is either real or imaginary. Since we want to find surface states it must be real (such that the wave function decays). Hence, as we already saw in section 4.1.2, λ_1, λ_2 must either both be real or complex conjugate partners. Furthermore, D_+D_- must be negative, which is the first criteria for the existence of surface states. If $A_1 = 0$ then $\lambda_1 + \lambda_2 = 0$, and they cannot both have a positive real part. Therefore, no surface states can exist, and our assumption that $A_1 \neq 0$ is justified.

If we insert the expression for λ_α from eq. (4.15), we get:

$$-\frac{A_1^2}{D_+D_-} = (\lambda_1+\lambda_2)^2 = \frac{-F+\sqrt{R}}{2D_+D_-} + \frac{-F-\sqrt{R}}{2D_+D_-} + \frac{\sqrt{F^2-R}}{|D_+D_-|} \tag{4.41}$$

$$\Leftrightarrow \sqrt{F^2-R} = \operatorname{sgn}(D_+D_-)(F-A_1^2) = A_1^2 - F$$

$$\Rightarrow F^2 - R = (A_1^2-F)^2 \tag{4.42}$$

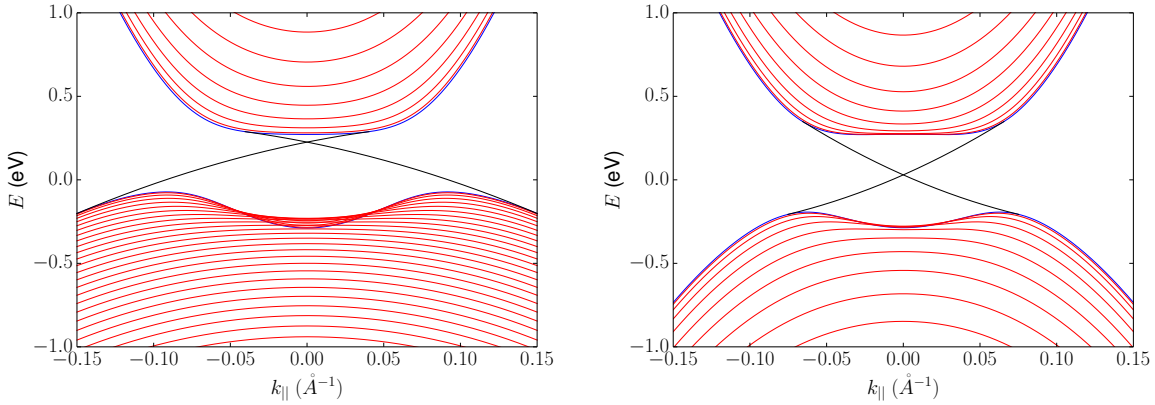
$$\Leftrightarrow 4D_+D_-((E-L_1)(E-L_2) - A_2^2k_{||}^2) = (D_+(E-L_1)+D_-(E-L_2))^2$$

$$\Leftrightarrow -4D_+D_-A_2^2k_{||}^2 = (D_+(E-L_1)-D_-(E-L_2))^2$$

$$\Leftrightarrow \pm\operatorname{sgn}(B_1)2\sqrt{B_1^2-D_1^2}|A_2|k_{||} = 2B_1(E-C-D_2k_{||}^2) + 2D_1(-M+B_2k_{||}^2) \tag{4.43}$$

$$\Leftrightarrow E = E_\pm(k_{||}) = C + \frac{MD_1}{B_1} \pm \sqrt{1-\frac{D_1^2}{B_1^2}}|A_2|k_{||} + \left(D_2 - \frac{D_1B_2}{B_1}\right)k_{||}^2. \tag{4.44}$$

This gives the spectrum of the surface states, if they exist. For small $k_{||}$ we get a linear Dirac dispersion, with a fermi velocity of $\frac{|A_2|}{\hbar}\sqrt{1-\frac{D_1^2}{B_1^2}}$. The $\operatorname{sgn}(B_1)$ is included, to ensure that the upper sign correspond to upper energy in both equations. Note that the arrow in eq. (4.42) goes only one way, since A_1^2-F could be negative. But if A_1^2-F is positive, then eq. (4.44) is equivalent to eq. (4.40). Therefore, given some in plane momentum $k_{||}$, a surface state exists with energy $E_\pm(k_{||})$ if and only if $A_1^2 - F(k_{||}, E_\pm(k_{||})) > 0$. A band of surface states can only end at point where $A_1^2 - F(k_{||}, E_\pm(k_{||})) = 0$, which by eq. (4.42) means that $F^2-R=0$. As we saw in section 4.1.2, this is exactly the $k_z = 0$ bulk spectrum, and we conclude that a band of surface states can only end by meeting the bulk bands. This is seen when plotting the surface and bulk spectrum, see figure 4.3.



(a) Parameters from [11] **(b)** Parameters from [12]

Figure 4.3: The spectrum of a semi-infinite topological insulator, including both bulk states (blue for $k_z = 0$ and red otherwise) and surface states (black).

The criteria for a surface state at is that $A_1^2 - F(k_{||}, E_\pm(k_{||})) > 0$. From the definition of

F in eq. (4.16), and using $E = E_a(k_{||})$ with $a = \pm$:

$$A_1^2 - F(k_{||}, E_a(k_{||})) = -D_+(E_a - L_1) - D_-(E_a - L_2) \tag{4.45}$$

$$= \frac{2B_2}{B_1} D_+ D_- k_{||}^2 - a2D_1|A_2|\sqrt{1 - \frac{D_1^2}{B_1^2}}k_{||} - \frac{2M}{B_1} D_+ D_- \tag{4.46}$$

Surface states exists, when this function is positive. Since we are working in a model, which is valid for small $k_{||}$ it is interesting to see whether we have surface states for $k_{||} = 0$:

$$-\frac{2M}{B_1} D_+ D_- > 0 \Leftrightarrow \frac{M}{B_1} > 0, \tag{4.47}$$

where we used the fact that $D_+ D_- < 0$. We conclude that to get two bands of surface states crossing in $k_{||} = 0$ we need the conditions:

$$D_+ D_- < 0 \text{ and } \frac{M}{B_1} > 0 \tag{4.48}$$

Note that the first condition is to get surface states at all, while the second is to get surface states at $k_{||} = 0$. Interestingly the first condition, equivalent to $|D_1| < |B_1|$, shows that the topologically non-trivial state can be broken by the particle-hole asymmetry. That can happen without the gap closing at the gamma point. If $|D_1| > |B_1|$ however, we do not have a global gap in the bulk spectrum, since both the conduction band and valence band energies diverges in the same direction, given by the sign of D_1 when $k_z \to \infty$.

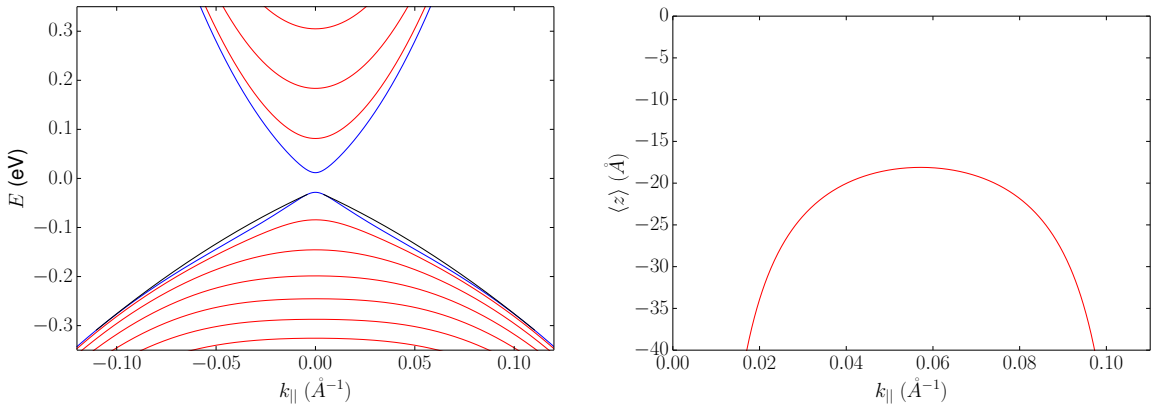
It is actually possible to have surface states without the second condition, but there will still be a gap. And we cannot get a Dirac-like spectrum since the crossing must occur at the gamma point. This happens, if we change the parameter M to 0.02 eV, while taking the rest of the parameters from [11]. Then $\frac{M}{B_1} < 0$, but we get a region of $k_{||}$ where there is surface states. This is actually a surprise since this is in the topologically trivial regime. These surface states go back into the lower band in both ends, as we see in figure 4.4a.

4.2.1 Experimental verification of surface states

The surface state spectrum can be measured, by angle-resolved photoemission spectroscopy (ARPES). This experimental technique is illustrated in 4.5. The sample is subjected to a beam of photons, and the electrons are emitted, due to the photoelectric effect. The energy and momentum of the emitted electrons are measured. By conservation of energy and momentum, this reveals the energies and momenta of the electronic states of the sample. This makes it possible to map out the dispersion relation of the electrons in a solid.

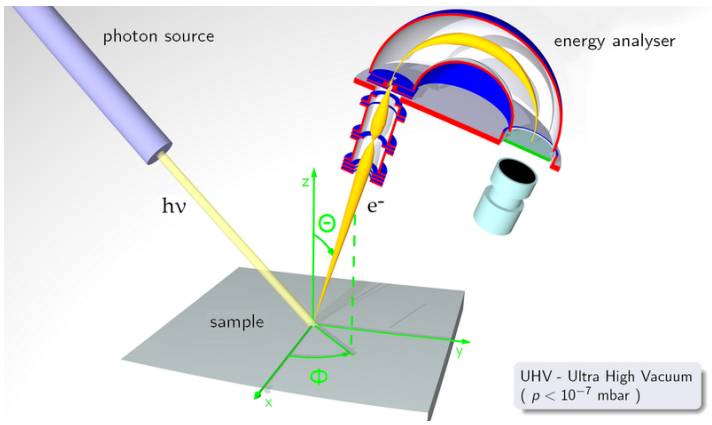
In figure 4.5b, we show the data from an ARPES experiment on a clean surface of Bi_2Se_3, from [21]. We see the qualitative agreement with the spectrum from the low energy model, shown in figure 4.3. In [7], a similar experiment was reported, and the fermi velocity was found to be approximately $5 \times 10^5\,\mathrm{ms}^{-1}$. The fermi velocity, in our model was given by:

$$\frac{|A_2|}{\hbar}\sqrt{1 - \frac{D_1^2}{B_1^2}} \tag{4.49}$$

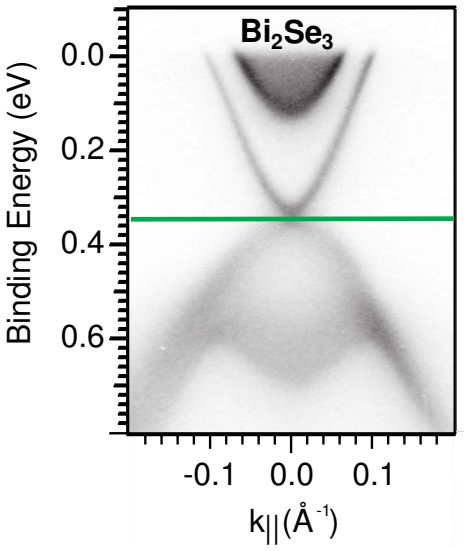


(a) **(b)**

Figure 4.4: (a) For $M = 0.02\,\text{eV}$, and the remaining parameters taken from [11], we see surface states (black lines) close to the bulk bands (blue for $k_z = 0$ and red lines), for a range of $k_{||}$. In (b) we see a plot of the expectation value of z, for the same parameters. The expectation value of z diverges in both ends, where the surface band approaches the bulk bands. Even though the surface band is close to the lower bulk band in energy, they are quite localized getting as close as $20\,\text{Å}$ from the surface.



(a) Adopted from [20]



(b) Adapted from [21]

Figure 4.5: (a) The experimental setup for an ARPES experiment. (b) The spectrum of the surface states found by ARPES found in [21]. Note that the energy is measured relative to the fermi level. The Fermi level level lies in the bulk conduction band, because naturally occuring Bi_2Se_3 is electron doped [7].

This gives $2.77 \times 10^5\,\mathrm{m\,s^{-1}}$ for the parameters from [11], and $6.17 \times 10^5\,\mathrm{m\,s^{-1}}$ for the parameters in [12]. This is not in exact agreement with the experimental results, but lies within an order of magnitude. Our model can not really be expected to give exact quantitative results, since it is only a second order approximation in $\mathbf{k}$. Even though we consider k_x and k_y to be small, we are investigating a phenomenon, which is localized in the z direction. Therefore, we should also expect some effects of higher order terms in k_z. However, this model does provide a clear physical picture, to understand the topological surface states.

4.3 Surface states at the gamma point

Now we will find the surface states at $k_{||} = 0$. To do this we will solve eq. (4.32), for the coefficients and then calculate the resulting wave functions. But since $k_{||} = 0$, the equation gets block diagonal. Each block corresponds to either spin up or down, since our basis spinor $\psi_{\alpha+1}$ only have spin up components, while $\psi_{\alpha+2}$ only have spin down components, for $k_{||} = 0$. This is due to the fact that the Hamiltonian gets block diagonal for $k_{||} = 0$. The secular eq. (4.35) now factorizes into:

$$\begin{aligned} 0 &= (\lambda_1 J_{2+} - \lambda_2 J_{1+})(\lambda_1 J_{2-} - \lambda_2 J_{1-}) \\ \Leftrightarrow \lambda_1 J_{2+} - \lambda_2 J_{1+} &= 0 \vee \lambda_1 J_{2-} - \lambda_2 J_{1-} = 0 \end{aligned} \tag{4.50}$$

Where each of the two equations correspond to one of the blocks in eq. (4.32). But since $J_{\alpha+} J_{\alpha-} = -A_1^2 \lambda_\alpha^2$ from eq. (4.27), the two equations are equivalent. Thus, we get a double degeneracy at the gamma point, with one spin up state and one spin down state. This is a result of the time reversal symmetry. For the spin up block we have the equation:

$$\begin{pmatrix} J_{1+} & J_{2+} \\ -iA_1\lambda_1 & -iA_1\lambda_2 \end{pmatrix} \begin{pmatrix} C_{1+1} \\ C_{2+1} \end{pmatrix} = 0 \tag{4.51}$$

If we choose $C_{1+1} = \frac{\sqrt{|D_-/B_1|}}{\sqrt{2}iA_1\lambda_1}$ we get $C_{2+1} = -\frac{\sqrt{|D_-/B_1|}}{\sqrt{2}iA_1\lambda_2}$. The vector multiplying $e^{\lambda_1 z}$ is simply:

$$C_{1+1}\psi_{1+1} = \begin{pmatrix} \frac{J_{1+}\sqrt{|D_-/B_1}}{\sqrt{2}iA_1\lambda_1} \\ -\frac{\sqrt{|D_-/B_1|}}{\sqrt{2}} \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} -\frac{i}{\sqrt{2}}\,\mathrm{sgn}(D_+A_1)\sqrt{\frac{|D_+|}{|B_1|}} \\ -\frac{1}{\sqrt{2}}\sqrt{\frac{|D_-|}{|B_1|}} \\ 0 \\ 0 \end{pmatrix}. \tag{4.52}$$

The second equality can be obtained using that $E = C + \frac{MD_1}{B_1}$ from eq. (4.44) and $\lambda_1 + \lambda_2 = \frac{|A_1|}{\sqrt{|D_+D_-|}}$ from eq. (4.40). The vector multiplying $e^{\lambda_2 z}$ must be the same, but with opposite overall sign to fulfill the boundary condition. The spin down state can obtained by take the

time reversal of the spin up state. The two eigenstates at the gamma point are then given by:

$$\Psi^{\uparrow}_{k_{||}=0}(z) = N \begin{pmatrix} -\frac{i}{\sqrt{2}}\operatorname{sgn}(D_+A_1)\sqrt{\frac{|D_+|}{|B_1|}} \\ -\frac{1}{\sqrt{2}}\sqrt{\frac{|D_-|}{|B_1|}} \\ 0 \\ 0 \end{pmatrix} (e^{\lambda_1 z} - e^{\lambda_2 z}) \tag{4.53}$$

$$\Psi^{\downarrow}_{k_{||}=0}(z) = N \begin{pmatrix} 0 \\ 0 \\ \frac{i}{\sqrt{2}}\operatorname{sgn}(D_+A_1)\sqrt{\frac{|D_+|}{|B_1|}} \\ -\frac{1}{\sqrt{2}}\sqrt{\frac{|D_-|}{|B_1|}} \end{pmatrix} (e^{\lambda_1 z} - e^{\lambda_2 z}) \tag{4.54}$$

where N is a normalization constant, given by:

$$1 = \int_{-\infty}^{0} (\Psi^{\sigma}_{k_{||}=0}(z))^{\dagger} \cdot \Psi^{\sigma}_{k_{||}=0}(z)\,\mathrm{d}z = N^2 \frac{1}{2|B_1|}(|D_+| + |D_-|) \int_{-\infty}^{0} |e^{\lambda_1 z} - e^{\lambda_2 z}|^2\,\mathrm{d}z \tag{4.55}$$

$$= N^2 \frac{1}{2|B_1|}(|D_+| + |D_-|) \int_{-\infty}^{0} e^{2\operatorname{Re}[\lambda_1]z} + e^{2\operatorname{Re}[\lambda_2]z} - 2\operatorname{Re}[e^{\lambda_1+\lambda_2^*)z}]\,\mathrm{d}z \tag{4.56}$$

$$= N^2 \frac{1}{2|B_1|}(|D_+| + |D_-|) \left(\frac{1}{2\operatorname{Re}[\lambda_1]} + \frac{1}{2\operatorname{Re}[\lambda_2]} - 2\operatorname{Re}[\frac{1}{\lambda_1 + \lambda_2^*}] \right) \tag{4.57}$$

The condition $D_+D_- < 0$ gives that $\operatorname{sgn}(B_1) = \operatorname{sgn}(D_+)$, which can be realized by considering the different options for the signs of D_1 and B_1. This gives that the spinor part of the wave function is already normalized, since:

$$\frac{1}{2|B_1|}(|D_+| + |D_-|) = \frac{\operatorname{sgn}(B_1)D_+ - \operatorname{sgn}(B_1)D_-}{2\operatorname{sgn}(B_1)B_1} = \frac{D_+ - D_-}{2B_1} = 1 \tag{4.58}$$

For the spatial part, either both λ's are real or complex conjugate partners, and in either case we get the normalization constant:

$$\begin{aligned} N &= \left(\frac{1}{2\lambda_1} + \frac{1}{2\lambda_2} - \frac{2}{\lambda_1 + \lambda_2} \right)^{-\frac{1}{2}} \quad \text{for } \lambda_1, \lambda_2 \in \mathbb{R} \\ N &= \left(\frac{1}{a} - \frac{a}{a^2 + b^2} \right)^{-\frac{1}{2}} \quad \text{for } \lambda_1 = \lambda_2^* = a + ib \end{aligned} \tag{4.59}$$

Choosing a different overall sign on the spin up state, the two states can be written:

$$\Psi^{\uparrow}_{k_{||}=0}(z) = \begin{pmatrix} \varphi(z) \\ 0 \end{pmatrix} \quad \Psi^{\downarrow}_{k_{||}=0} = \begin{pmatrix} 0 \\ \tau_z\varphi(z) \end{pmatrix} \tag{4.60}$$

where

$$\varphi(z) = N \begin{pmatrix} \frac{i}{\sqrt{2}}\operatorname{sgn}(D_+A_1)\sqrt{\frac{|D_+|}{|B_1|}} \\ \frac{1}{\sqrt{2}}\sqrt{\frac{|D_-|}{|B_1|}} \end{pmatrix} (e^{\lambda_1 z} - e^{\lambda_2 z}) \tag{4.61}$$

We will use these states to construct a 2D model, describing the surface states. But before doing this we will calculate the surface states at general in-plane momentum in the 3D model, such that we can compare the results to the 2D model.

4.4 Surface states at $k_{||} \neq 0$

Now we will find the wave functions for the surface states at a general non-zero in-plane momentum $k_{||}$. To do this we need to find the coefficients from eq. (4.32), and calculate the full wave function from 4.28. We can choose one of the coefficients as we please, so lets choose

$$C_{1+1} = \frac{\sqrt{|D_-/B_1|}}{i2A_1\lambda_1},$$

The second and third row of eq. (4.32) gives:

$$C_{2+1} = -\frac{\lambda_1}{\lambda_2}C_{1+1} \tag{4.62}$$

$$C_{2+2} = -\frac{\lambda_1}{\lambda_2}C_{1+2} \tag{4.63}$$

if we insert these into the first row of eq. (4.32), we get:

$$C_{1+2} = -\frac{J_{1+} - \frac{\lambda_1}{\lambda_2}J_{2+}}{A_2k_-(1-\frac{\lambda_1}{\lambda_2})}\frac{\sqrt{|D_-/B_1|}}{i2A_1\lambda_1} \tag{4.64}$$

$$= \frac{D_+\lambda_1\lambda_2 - E + L_2}{2iA_1A_2\lambda_1k_-}\sqrt{\frac{|D_-|}{|B_1|}} \tag{4.65}$$

And we now have the coefficients:

$$C_{1+1} = \frac{1}{2iA_1\lambda_1}\sqrt{\frac{|D_-|}{|B_1|}} \tag{4.66}$$

$$C_{1+2} = \frac{D_+\lambda_1\lambda_2 - E + L_2}{2iA_1A_2\lambda_1k_-}\sqrt{\frac{|D_-|}{|B_1|}} \tag{4.67}$$

$$C_{2+1} = -\frac{1}{2iA_1\lambda_2}\sqrt{\frac{|D_-|}{|B_1|}} \tag{4.68}$$

$$C_{2+2} = -\frac{D_+\lambda_1\lambda_2 - E + L_2}{2iA_1A_2\lambda_2k_-}\sqrt{\frac{|D_-|}{|B_1|}} \tag{4.69}$$

the vector multiplying $e^{\lambda_1 z}$ is now given by:

$$C_{1+1}\psi_{1+1} + C_{1+2}\psi_{1+2} = \frac{1}{2iA_1\lambda_1}\sqrt{\frac{|D_-|}{|B_1|}}\begin{pmatrix} D_+\lambda_1^2 - L_2 + E \\ -iA_1\lambda_1 \\ 0 \\ A_2k_+ \end{pmatrix} \tag{4.70}$$

$$+\frac{D_+\lambda_1\lambda_2 - E + L_2}{2iA_1A_2\lambda_1k_-}\sqrt{\frac{|D_-|}{|B_1|}}\begin{pmatrix} A_2k_- \\ 0 \\ iA_1\lambda_1 \\ D_-\lambda_1^2 - L_1 + E \end{pmatrix} \tag{4.71}$$

To calculate the actual surface state, we need to insert the energy $E_\pm(k_{||})$ from eq. (4.44) into this vector. However, sometimes it will be simpler to use one of the equations from the derivation of eq. (4.44). The first component is:

$$\frac{1}{2iA_1\lambda_1}\sqrt{\frac{|D_-|}{|B_1|}}(D_+\lambda_1^2 - L_2 + E + D_+\lambda_1\lambda_2 - E + L_2) \tag{4.72}$$

$$= \frac{1}{2iA_1}\sqrt{\frac{|D_-|}{|B_1|}}D_+(\lambda_1 + \lambda_2) = \frac{-i}{2}\operatorname{sgn}(A_1D_+)\sqrt{\frac{|D_+|}{|B_1|}} \tag{4.73}$$

where we used that $(\lambda_1 + \lambda_2) = \frac{|A_1|}{\sqrt{|D_+D_-|}}$ from eq. (4.40). The second component is trivially $-\frac{1}{2}\sqrt{\frac{|D_-|}{|B_1|}}$. To calculate the third component, we will first use that $\lambda_1\lambda_2 = \frac{\sqrt{F^2-R}}{2|D_+D_-|}$ from eq. (4.15). Then we will use both $\sqrt{F^2-R} = \operatorname{sgn}(D_+D_-)(F-A_1^2)$ and $D_+(E-L_1)-D_-(E-L_2) = \pm 2\operatorname{sgn}(B_1)|A_2|k_{||}\sqrt{|D_+D_-|}$, from the derivation of the surface spectrum, eq. (4.44). This gives the third component:

$$\frac{D_+\lambda_1\lambda_2 - E + L_2}{2A_2k_-}\sqrt{\frac{|D_-|}{|B_1|}} = -\frac{\operatorname{sgn}(D_+D_-)\frac{\sqrt{F^2-R}}{2D_-} - E + L_2}{2A_2k_-}\sqrt{\frac{|D_-|}{|B_1|}} \tag{4.74}$$

$$= \frac{(F - A_1^2) - 2D_-(E - L_2)}{4A_2k_-D_-}\sqrt{\frac{|D_-|}{|B_1|}} \tag{4.75}$$

$$= \frac{D_+(E - L_1) - D_-(E - L_2)}{4A_2k_-D_-}\sqrt{\frac{|D_-|}{|B_1|}} \tag{4.76}$$

$$= \pm\operatorname{sgn}(B_1)\frac{2|A_2|k_{||}\sqrt{|D_+D_-|}}{4A_2k_-D_-}\sqrt{\frac{|D_-|}{|B_1|}} \tag{4.77}$$

$$= \pm\operatorname{sgn}(D_-A_2B_1)\frac{k_+}{2k_{||}}\sqrt{\frac{|D_+|}{|B_1|}} \tag{4.78}$$

The fourth component is:

$$\begin{aligned}
&\frac{1}{2iA_1\lambda_1}\sqrt{\frac{|D_-|}{|B_1|}}\left(A_2k_+ + \frac{(D_+\lambda_1\lambda_2 - E + L_2)(D_-\lambda_1^2 - L_1 + E)}{A_2k_-}\right) \\
&= \frac{1}{2iA_1\lambda_1A_2k_-}\sqrt{\frac{|D_-|}{|B_1|}} \\
&\quad \times \big(D_+D_-\lambda_1^3\lambda_2 + D_+\lambda_1\lambda_2(E-L_1) - D_-\lambda_1^2(E-L_2) \\
&\quad + A_2^2k_{||}^2 - (E-L_1)(E-L_2)\big) \qquad (4.79) \\
&= \frac{1}{2iA_1A_2k_-\lambda_1}\sqrt{\frac{|D_-|}{|B_1|}} \\
&\quad \times \big(D_+D_-\lambda_1^3\lambda_2 + D_+\lambda_1\lambda_2(E-L_1) - D_-\lambda_1^2(E-L_2) - D_+D_-\lambda_1^2\lambda_2^2\big) \qquad (4.80) \\
&= \frac{1}{2iA_1A_2k_-(\lambda_1+\lambda_2)}\sqrt{\frac{|D_-|}{|B_1|}}\big(D_+D_-\lambda_1\lambda_2(\lambda_1^2-\lambda_2^2) + D_+\lambda_2^2(E-L_1) \\
&\quad - D_-\lambda_1^2(E-L_2) + \lambda_1\lambda_2(D_+(E-L_1) - D_-(E-L_2))\big) \qquad (4.81)
\end{aligned}$$

Now we will insert the expression from eq. (4.15) for the squared λ's. Again we will use both $\sqrt{F^2-R} = \operatorname{sgn}(D_+D_-)(F-A_1^2)$ and $D_+(E-L_1)-D_-(E-L_2) = \pm 2\operatorname{sgn}(B_1)|A_2|k_{||}\sqrt{|D_+D_-|}$, from the derivation of the surface spectrum, eq. (4.44). The product then becomes $\lambda_1\lambda_2 = \frac{\sqrt{F^2-R}}{2|D_+D_-|} = \frac{F-A_1^2}{2D_+D_-}$. Inserting all this in eq. (4.81), we get the fourth component:

$$\begin{aligned}
&\frac{1}{4iA_1A_2k_-(\lambda_1+\lambda_2)D_+D_-}\sqrt{\frac{|D_-|}{|B_1|}}\big(\sqrt{R}(D_+(E-L_1) + D_-(E-L_2)) \qquad (4.82) \\
&- F(D_+(E-L_1) - D_-(E-L_2)) - \sqrt{R}(D_+(E-L_1) + D_-(E-L_2)) \\
&+ (F-A_1^2)(D_+(E-L_1) - D_-(E-L_2))\big) \\
&= \frac{-A_1^2(D_+(E-L_1) - D_-(E-L_2))}{4iA_1A_2k_-(\lambda_1+\lambda_2)D_+D_-}\sqrt{\frac{|D_-|}{|B_1|}} \\
&= \frac{\mp\operatorname{sgn}(B_1)2A_1|A_2|k_{||}\sqrt{|D_+D_-|}}{4iA_2k_-\frac{|A_1|}{\sqrt{|D_+D_-|}}D_+D_-}\sqrt{\frac{|D_-|}{|B_1|}} \\
&= \mp\operatorname{sgn}(A_1A_2B_1)\operatorname{sgn}(D_+D_-)\frac{k_+}{2ik_{||}}\sqrt{\frac{|D_-|}{|B_1|}} = \mp\operatorname{sgn}(A_1A_2B_1)\frac{ik_+}{2k}\sqrt{\frac{|D_-|}{|B_1|}} \qquad (4.83)
\end{aligned}$$

Finally we arrived at a simplified expression for the vector multiplying $e^{\lambda_1 z}$:

$$\begin{pmatrix} \frac{-i}{2}\operatorname{sgn}(A_1D_+)\sqrt{\frac{|D_+|}{|B_1|}} \\ -\frac{1}{2}\sqrt{\frac{|D_-|}{|B_1|}} \\ \pm\operatorname{sgn}(D_-A_2B_1)\frac{k_+}{2k_{||}}\sqrt{\frac{|D_+|}{|B_1|}} \\ \mp\operatorname{sgn}(A_1A_2B_1)\frac{ik_+}{2k_{||}}\sqrt{\frac{|D_-|}{|B_1|}} \end{pmatrix} \qquad (4.84)$$

where the upper sign corresponds to the energy E_+ and the lower to E_-. The vector multi-

plying $e^{\lambda_2 z}$ is clearly the same but with the opposite overall sign, since the wave function must vanish at the surface. This gives the total wave functions:

$$\Psi^{\pm}_{k_x,k_y}(z) = N \begin{pmatrix} \frac{-i}{2}\,\mathrm{sgn}(A_1B_1)\sqrt{\frac{|D_+|}{|B_1|}} \\ -\frac{1}{2}\sqrt{\frac{|D_-|}{|B_1|}} \\ \mp\,\mathrm{sgn}(A_2)\frac{k_+}{2k_{||}}\sqrt{\frac{|D_+|}{|B_1|}} \\ \mp\,\mathrm{sgn}(A_1A_2B_1)\frac{ik_+}{2k_{||}}\sqrt{\frac{|D_-|}{|B_1|}} \end{pmatrix} (e^{\lambda_1(k_{||},E_\pm(k_{||}))z} - e^{\lambda_2(k_{||},E_\pm(k_{||}))z}). \tag{4.85}$$

We see that the resulting wave spinor part of the wave function is a superposition of the spinors for the surface states at the gamma point, but where the the two different spin parts have a relative phase of $\frac{k_+}{k_{||}}$. Therefore, N is again given by eq. (4.59). We can insert the energy from eq. (4.44), to express λ_1 and λ_2 as functions of $k_{||}$ only. We have:

$$\lambda_\alpha(k_{||}, E_\pm(k_{||})) = \sqrt{\frac{-F + (-1)^{\alpha-1}\sqrt{R}}{2D_+D_-}}, \tag{4.86}$$

where the F and R after some algebra, can be simplified to:

$$F = A_1^2 + 2\frac{D_+D_-}{B_1}(M - B_2k_{||}^2) \pm 2D_1|A_2|\sqrt{1 - \frac{D_1^2}{B_1^2}}k_{||}, \tag{4.87}$$

$$R = A_1^4 + \frac{4A_1^2}{B_1}D_+D_-(M - B_2k_{||}^2) \pm 4A_1D_1|A_2|\sqrt{1 - \frac{D_1^2}{B_1^2}}k_{||}. \tag{4.88}$$

4.4.1 Spatial structure

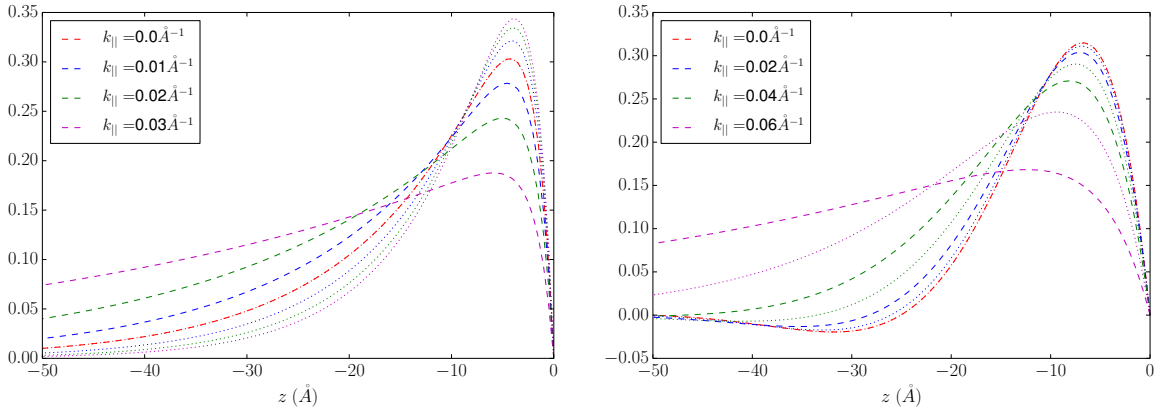
Now we will discuss the spatial distribution of the surface states in the z direction. This is entirely determined by the values of λ_1 and λ_2, which are given in terms of $k_{||}$, in eq. (4.86). Because of the particle-hole asymmetry, λ_1 and λ_2 also depend on whether which part of the Dirac cone we are on. For a particle-hole symmetric model $D_1 = 0$, and the terms with a $\pm$ in eq. (4.87) and (4.88) drop out. At $k_{||} = 0$ we directly see that the spatial distribution of the surface states are equal for E_+ and E_-, since $E_+ = E_-$. This is can also be realized from time reversal symmetry; at $k_{||}$, the two surface states are related by the time reversal operator, which does not change the spatial part. Qualitatively, there is two different options for the spatial part of the surface states. Either we have that λ_1 and λ_2 are real, or they are complex conjugate partners. If they are complex conjugate partners $\lambda_1 = \lambda_2^* = a + ib$, then the spatial part of the wave function can be written:

$$e^{\lambda_1 z} - e^{\lambda_2 z} = e^{az}(e^{ib} - e^{-ib}) = 2ie^{az}\sin(bz), \tag{4.89}$$

and the wave function is simply a sine function multiplied by an exponential decay.

For the parameters from [11], the surface bands are entirely in the region where $\lambda_1, \lambda_2 \in \mathbb{R}$, and as we see on figure 4.6a, the wave functions show a simple decay, without any oscillation. If we instead use the parameters from [12], the wave function oscillates for small $k_{||}$. For larger $k_{||}$ the imaginary part b decreases, and the wave length of the oscillation increases, and

eventually diverges when λ_1 and λ_2 becomes real. On figure 4.6b, we see the wave length increases with $k_{||}$, and at $k_{||} = 0.06\,\text{Å}^{-1}$, we see a simple decay. In chapter 5, we will see how this oscillatory behavior is of great importance, when considering a finite topological insulator.



(a) Parameters from [11]. Note how the lower band gets more localized while the lower goes into the bulk for this range of $k_{||}$. For larger $k_{||}$ wave functions of both bands goes into the bulk, as seen on figure 4.7.

(b) Parameters from [12]. The wave length of the oscillation increases with $k_{||}$ and for $k_{||} = 0.06\,\text{Å}^{-1}$ we see a decay without any oscillation.

Figure 4.6: Wave functions for a semi-infinite topological insulator. The dashed lines are from the upper band E_+ while the dotted ones are from the upper band E_-. The color indicates the value of the in plane momentum $k_{||}$ (note that the values of $k_{||}$ are different for the two plots). The spatial wave functions for the two bands are equal for $k_{||} = 0$ since in this point they are Kramers partners.

It is also interesting to see how localized the surface states are. As we see in figure 4.6, this changes with $k_{||}$. One way of quantifying this is to calculate the expectation value of the z:

$$\langle z \rangle = N^2 \int_{-\infty}^{0} z(\Psi^{\pm}_{k_x,k_y})^{\dagger}\Psi^{\pm}_{k_x,k_y}\,\mathrm{d}z = N^2 \int_{-\infty}^{0} ze^{2\,\mathrm{Re}[\lambda_1]z} + ze^{2\,\mathrm{Re}[\lambda_2]z} - 2z\,\mathrm{Re}[e^{\lambda_1+\lambda_2^*)z}]\,\mathrm{d}z \tag{4.90}$$

$$= N^2\left(-\frac{1}{(2\,\mathrm{Re}[\lambda_1])^2} - \frac{1}{(2\,\mathrm{Re}[\lambda_2])^2} + 2\,\mathrm{Re}\left[\frac{1}{(\lambda_1+\lambda_2^*)^2}\right]\right) \tag{4.91}$$

$$\tag{4.92}$$

Note that λ_1 and λ_2 are given by eq. (4.86), we have just dropped the arguments for convenience. Again we can simplify by looking at either $\lambda_1, \lambda_2 \in \mathbb{R}$ or $\lambda_1 = \lambda_2^* = a + ib$. For the real case we get:

$$\langle z \rangle = \frac{-\frac{1}{4\lambda_1^2} - \frac{1}{4\lambda_1^2} + \frac{2}{(\lambda_1+\lambda_2)^2}}{\frac{1}{2\lambda_1} + \frac{1}{2\lambda_2} - \frac{2}{\lambda_1+\lambda_2}} = -\frac{1}{2\lambda_1} - \frac{1}{2\lambda_2} - \frac{1}{\lambda_1+\lambda_2} \tag{4.93}$$

And for the complex case:

$$\langle z \rangle = \frac{-\frac{1}{2a^2} + \frac{a^2-b^2}{2(a^2+b^2)^2}}{\frac{1}{a} - \frac{a}{a^2+b^2}} = -\frac{1}{2a} - \frac{a}{a^2+b^2} \tag{4.94}$$

In figure 4.7 the expectation value is plotted as a function of the in plane momentum. By comparing with the spectra in figure 4.3, we see that the expectation value diverges, exactly when the surface band touch the $k_z = 0$ bulk band. For $k_{||} = 0$ the expectation value is bigger than -10 Å, for both sets of parameters, which is within the first quintuple layer. For the parameters from [11], the lower surface state with energy E_-, actually gets more localized as $k_{||}$ increases, before going into the bulk.

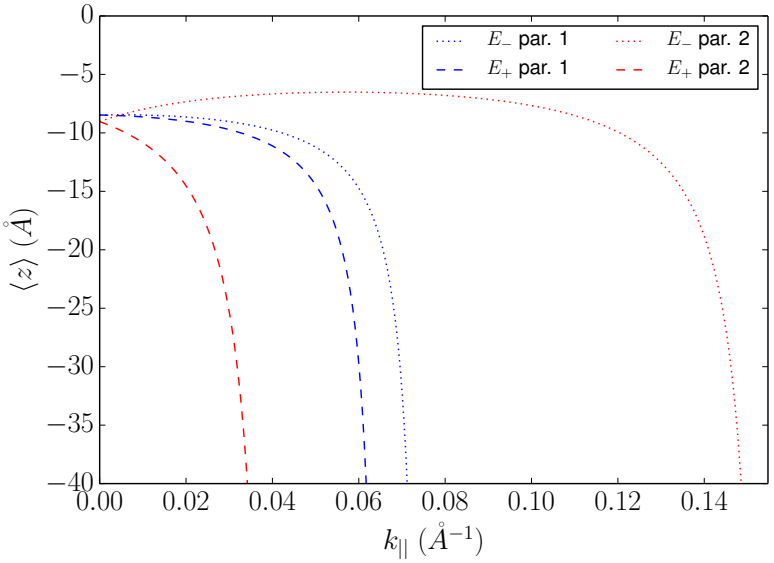


Figure 4.7: Expectation value of z for the surface states of a semi infinite insulator. The red lines are from parameters from [11], while the blue ones are from the parameters of [12].

4.4.2 Spin structure

To investigate the spin structure of these surface states we will transform back to the original basis, related by the unitary matrix U_1 given in eq. (3.45). In that basis the σ matrices represents spin directly:

$$U_1^\dagger \Psi^\pm_{k_x,k_y}(z) = N \begin{pmatrix} \frac{-i}{2}\operatorname{sgn}(A_1 B_1)\sqrt{\frac{|D_+|}{|B_1|}} \\ -\frac{i}{2}\sqrt{\frac{|D_-|}{|B_1|}} \\ \mp \operatorname{sgn}(A_2)\frac{k_+}{2k_{||}}\sqrt{\frac{|D_+|}{|B_1|}} \\ \mp \operatorname{sgn}(A_1 A_2 B_1)\frac{k_+}{2k_{||}}\sqrt{\frac{|D_-|}{|B_1|}} \end{pmatrix} (e^{\lambda_1 z} - e^{\lambda_2 z}) \tag{4.95}$$

$$= N \begin{pmatrix} \frac{i}{\sqrt{2}} \\ \pm \operatorname{sgn}(A_1 A_2 B_1)\frac{k_+}{\sqrt{2}k_{||}} \end{pmatrix}_{\text{spin}} \otimes \begin{pmatrix} -\operatorname{sgn}(A_1 B_1)\sqrt{\frac{|D_+|}{2|B_1|}} \\ -\sqrt{\frac{|D_-|}{2|B_1|}} \end{pmatrix}_{\text{orbital}} (e^{\lambda_1 z} - e^{\lambda_2 z}) \tag{4.96}$$

We see that the spinor splits into a spin part and an orbital part, and that the spin part is independent of the parameters of the model except for the sign of the parameters A_1, A_2 and

B_1. The expectation values of the spin operators are:

$$\langle S_z \rangle_{E_\pm} = \frac{1}{2} \begin{pmatrix} -\frac{i}{\sqrt{2}} \\ \pm \operatorname{sgn}(A_1 A_2 B_1) \frac{k_-}{\sqrt{2} k_{||}} \end{pmatrix} \cdot \begin{pmatrix} \frac{i}{\sqrt{2}} \\ \mp \operatorname{sgn}(A_1 A_2 B_1) \frac{k_+}{\sqrt{2} k_{||}} \end{pmatrix} = 0 \tag{4.97}$$

$$\langle S_x \rangle_{E_\pm} = \frac{1}{2} \begin{pmatrix} -\frac{i}{\sqrt{2}} \\ \pm \operatorname{sgn}(A_1 A_2 B_1) \frac{k_-}{\sqrt{2} k_{||}} \end{pmatrix} \cdot \begin{pmatrix} \pm \operatorname{sgn}(A_1 A_2 B_1) \frac{k_+}{\sqrt{2} k_{||}} \\ \frac{i}{\sqrt{2}} \end{pmatrix} = \pm \operatorname{sgn}(A_1 A_2 B_1) \frac{k_y}{2 k_{||}} \tag{4.98}$$

$$\langle S_y \rangle_{E_\pm} = \frac{1}{2} \begin{pmatrix} -\frac{i}{\sqrt{2}} \\ \pm \operatorname{sgn}(A_1 A_2 B_1) \frac{k_-}{\sqrt{2} k_{||}} \end{pmatrix} \cdot \begin{pmatrix} \mp i \operatorname{sgn}(A_1 A_2 B_1) \frac{k_+}{\sqrt{2} k_{||}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} = \mp \operatorname{sgn}(A_1 A_2 B_1) \frac{k_x}{2 k_{||}} \tag{4.99}$$

We see that spin and momentum are locked such that the spin is always perpendicular to the momentum, and the only dependence on the parameters of the model is the sign of the vorticity of the spin, which is determined by the sign of $A_1 A_2 B_1$. For the parameters from [11], the product $A_1 A_2 B_1$ is negative, and we get the spin structure given in figure 4.8. Note that in [11], the spin structure is calculated as well, using an effective 2D model, giving the opposite vorticity of the spin. The reason is that they consider a system where the topological insulator fills the $z > 0$ half-space. The two systems are related by the inversion operator, which takes $k_x, k_y \to -k_x, -k_y$, but does not affect the spin. Therefore, the vorticity of the spin structure is reversed. For the parameters in [12], $A_1 A_2 B_1$ is actually positive, and the spin structure has opposite vorticity. These parameters, however, were obtained by fitting the bulk spectrum of the model to the spectrum from ab initio calculations. The bulk spectrum does not depend on the signs of the parameters, A_1, A_2 and B_1 and therefore, this method gives a sign ambiguity. Hence, the vorticity of the spin structure is not well determined by this method only. In [11] the parameters are calculated using $k\cdot$ perturbation theory, where the parameters are given by matrix elements of the momentum operator between the bulk states at the gamma point. Therefore, these parameters should be unambiguous.

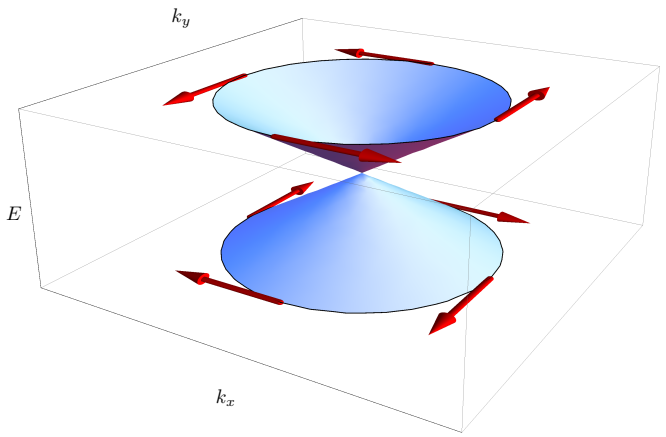


Figure 4.8: Close up of the Dirac cone of surface states. The spin is shown by the red arrows. The spin lies within the xy plane and rotates around the gamma point, in a counter-clockwise fashion above the Dirac crossing, and clockwise below the crossing. This is the case for the parameters from [11], while all the spins are all opposite using the parameters from [12].

4.5 2D model for surface states

Using the two surface states at the gamma point as basis states, we will now construct a 2D model, describing the surface electrons. We will do this including only terms to second order in $k_{||}$, since then we can compare it to the analytical solution in 4.4. This can be thought of as doing degenerate perturbation theory, in the in plane momentum. We split the Hamiltonian into one term independent of k_x and k_y and the perturbation term depending on k_x and k_y:

$$H(k_x, k_y, -i\partial_z) = H_0(-i\partial_z) + \Delta H(k_x, k_y) \tag{4.100}$$

where $H_0(-i\partial_z) = H(k_x = 0, k_y = 0, -i\partial_z)$ and

$$\Delta H(k_x, k_y) = D_2 k_{||}^2 - B_2 k_{||} \sigma_0 \otimes \tau_z + A_2(k_x \sigma_x + k_y \sigma_y) \otimes \tau_x \tag{4.101}$$

To get the effective 2D Hamiltonian, we need to calculate the matrix elements between our basis states $\Psi^{\uparrow}, \Psi^{\downarrow}$ (we drop the $k_{||} = 0$ subscript here for convenience) with the perturbation $\Delta H(k_x, k_y)$. Since there is no operator in $\Delta H(k_x, k_y)$ operating in real space, we do not have to worry about the real space part of the wave function. Dropping the spatial part, our basis states are:

$$\Psi^{\uparrow} = \begin{pmatrix} \varphi \\ 0 \end{pmatrix} \quad \Psi^{\downarrow} = \begin{pmatrix} 0 \\ \tau_z \varphi \end{pmatrix} \tag{4.102}$$

where

$$\varphi = \begin{pmatrix} \frac{i}{\sqrt{2}} \operatorname{sgn}(D_+ A_1) \sqrt{\frac{|D_+|}{|B_1|}} \\ \frac{1}{\sqrt{2}} \sqrt{\frac{|D_-|}{|B_1|}} \end{pmatrix} \tag{4.103}$$

The elements of our 2D effective Hamiltonian are then given by

$$H^{2D}_{\sigma\sigma'} = \langle \Psi^{\sigma} | \Delta H | \Psi^{\sigma'} \rangle \tag{4.104}$$

Thus the first term is just gives a constant. For the second term we need to calculate the matrix elements of $\sigma_0 \otimes \tau_z$. This term is diagonal in spin space and thus only give diagonal terms in our effective hamiltionian. These matrix elements give:

$$\langle \Psi^{\uparrow} | \sigma_0 \otimes \tau_z | \Psi^{\uparrow} \rangle = \langle \varphi | \tau_z | \varphi \rangle = \frac{|D_+| - |D_-|}{2|B_1|} = \frac{\operatorname{sgn}(B_1) D_+ + \operatorname{sgn}(B_1) D_-}{2 \operatorname{sgn}(B_1) B_1} = \frac{D_1}{B_1} \tag{4.105}$$

$$\langle \Psi^{\downarrow} | \sigma_0 \otimes \tau_z | \Psi^{\downarrow} \rangle = \langle \varphi | \tau_z \tau_z \tau_z | \varphi \rangle = \langle \varphi | \tau_z | \varphi \rangle = \frac{D_1}{B_1} \tag{4.106}$$

For the last term the spin is flipped, and we get only off diagonal terms. We will just calculate one of them and get the other by taking the complex conjugate:

$$\langle\Psi^{\uparrow}|A_2\begin{pmatrix}0 & k_-\\ k_+ & 0\end{pmatrix}\otimes\tau_x|\Psi^{\downarrow}\rangle = A_2k_-\langle\varphi|\tau_x\tau_z|\varphi\rangle = -iA_2k_-\langle\varphi|\tau_y|\varphi\rangle = iA_2\,\mathrm{sgn}(B_1A_1)\sqrt{1-\frac{D_1^2}{B_1^2}} \tag{4.107}$$

Putting this together, and using that the energy of the unperturbed states is $E = C + \frac{MD_1}{B_1}$ we get the 2D Hamiltonian:

$$H^{2D} = C + \frac{MD_1}{B_1} + A_2\,\mathrm{sgn}(B_1A_1)\sqrt{1-\frac{D_1^2}{B_1^2}}(k_y\sigma_x - k_x\sigma_y) + \left(D_2 - \frac{B_2D_1}{B_1}\right)k_{||} \tag{4.108}$$

The linear term is the Dirac-like Hamiltonian from eq. (1.2), giving rise to spin-momentum locking. The sign of the vorticity of the spin-momentum locking is determined by the sign of $A_1A_2B_1$, exactly like the we saw using the 3D model in section 4.4.2. In addition we also get a second order term, which only changes the spectrum, giving a curvature of the dispersion relation, and not the spin structure of the surface states. This Hamiltonian can easily be solved giving the same spectrum, as the full solution to the 3D model, eq. (4.44).

4.6 Local density of states

Now we will calculate the local density of states. This will in principle not give any new information, but will sum up the information given by the wave functions. From an experimental point of view this is interesting, since you do not measure single wave functions. By doing for example an STM measurement, you probe the local density of states. This can be computed directly from the wave functions, where we now include the x and y dependence from eq. (4.10):

$$\rho(z,\omega) = \sum_{k_x,k_y,\alpha}\frac{1}{A}|\Psi^{\alpha}_{k_x,k_y}(z)|^2\delta(\omega - E_\alpha(k_{||})) \tag{4.109}$$

where $\alpha = \pm$ denotes if we are on the upper or lower part of the Dirac cone, and A is the area in the x, y plane, which is included to normalize the wave functions. But we will take the limit $A \to \infty$, which converts the sums over k_x and k_y to integrals as described in Appendix A of [22]:

$$\rho(z,\omega) = \frac{1}{(2\pi)^2}\sum_{\alpha}\int_{\mathrm{FBZ}}\mathrm{d}k_x\mathrm{d}k_y|\Psi_{k_x,k_y,\alpha}(z)|^2\delta(\omega - E_\alpha(k_{||})) \tag{4.110}$$

Since the Hamiltonian is invariant under any rotation around the z axis, $|\Psi_{k_x,k_y,\alpha}(z)|^2 = |\Psi_{k_{||},0,\alpha}(z)|^2$. Therefore, we change to polar coordinates in the momentum integrals and inte-

grate out the angle:

$$\rho(z,\omega) = \frac{1}{(2\pi)^2}\sum_\alpha \int \mathrm{d}k_{||}k_{||}\int \mathrm{d}\theta |\Psi_{k_{||},0,\alpha}(z)|^2\delta(\omega - E_\alpha(k_{||})) \tag{4.111}$$

$$= \frac{1}{2\pi}\sum_\alpha \int \mathrm{d}k_{||}k_{||}|\Psi_{k_{||},0,\alpha}(z)|^2\delta(\omega - E_\alpha(k_{||})) \tag{4.112}$$

$$= \frac{1}{2\pi}\sum_\alpha \int \mathrm{d}k_{||}k_{||}|\Psi_{k_{||},0,\alpha}(z)|^2\sum_{k_0}\frac{\delta(k_{||}-k_0))}{\frac{\mathrm{d}E_\alpha}{\mathrm{d}k_{||}}(k_0)} \tag{4.113}$$

where k_0 are the solutions to the equation $E_\alpha(k_0) = \omega$.

$$\omega = E_\alpha = ak_0^2 + \alpha b k_0 + c \tag{4.114}$$

$$k_0 = \frac{-\alpha b \pm \sqrt{b^2 - 4a(c-\omega)}}{2a} \tag{4.115}$$

where $a = D_2 - \frac{D_1B_2}{B_1}$, $b = \sqrt{1-\frac{D_1^2}{B_1^2}}|A_2|$ and $c = C + \frac{MD_1}{B_1}$. $\alpha = \pm$ denotes whether we are on the upper or lower part of the Dirac cone. This equation has two solutions, but we only want $k_0 > 0$ solutions since it is the magnitude of the in plane momentum. We also see on figure 4.3 the surface the surface bands go into the bulk before the parabola reaches its extremum. Hence, if both solutions are positive, we only want the smallest positive solution. Hence we have a unique solution for each ω. By considering all 4 combinations of the signs of a and $c-\omega$, we see that the smallest positive solution always can be written:

$$k_0(\omega) = \mathrm{sgn}(c-\omega)\frac{b - \sqrt{b^2-4a(c-\omega)}}{2a} \tag{4.116}$$

and the derivative of the energy in this point is:

$$\frac{\mathrm{d}E}{\mathrm{d}k_{||}}(k_0) = \mathrm{sgn}(\omega - c)b + 2ak_0 \tag{4.117}$$

And we get the local density of states:

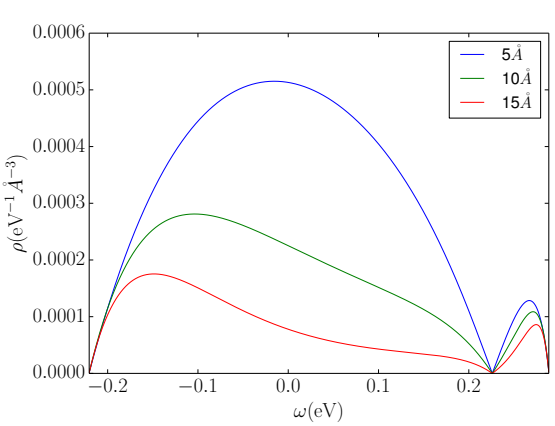
$$\rho(z,\omega) = \frac{1}{2\pi}k_0(\omega)\frac{|\Psi_{k_0,0,\mathrm{sgn}(\omega-c)}(z)|^2}{|\mathrm{sgn}(\omega-c)b + 2ak_0(\omega)|} \tag{4.118}$$

And if we write this out using the specific form of the wave function, we get:

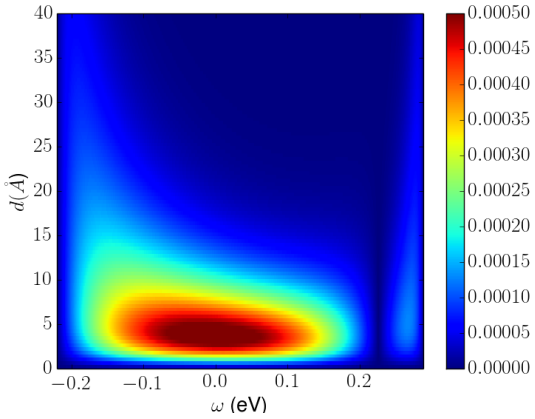
$$\rho(z,\omega) = \frac{1}{2\pi}k_0(\omega)\frac{N^2|e^{\lambda_1 z} - e^{\lambda_1 z}|^2}{|\mathrm{sgn}(\omega-c)b + 2ak_0(\omega)|} \tag{4.119}$$

where λ_1 and λ_2 are given by eq. (4.15) with $E = \omega$ and $k = k_0(\omega)$, and N given by eq. (4.59).

The density of states is zero at the Dirac point. The reason is that here we only get a contribution from a single pair of states, whereas for all other energies, there is a circle of contributing states in momentum space. The linear dependence of the DOS close to the Dirac point is a characteristic of a Dirac Hamiltonian. When the energy moves away from the Dirac



(a) The local density of states for specific distances from the surface, as a function of the energy ω.



(b) The local density of states, as a function of energy ω and distance d from the surface for the semi infinite system.

Figure 4.9: The local density of states for the surface states of a semi-infinite topological insulator. The Dirac point is seen at $\omega = 0.23$ Å where the density becomes zero. We see how as we go away from the Dirac point the density goes further into the bulk, and eventually goes to zero.

point, the wave function spreads further into the material, and as a result we see the LDOS decreasing and eventually becoming zero, when the surface state becomes a bulk state.

Surface states in a thin film of Bi_2Se_3

In this chapter, we will investigate a finite topological insulator. We consider a system, which is infinite in the x, and y directions but, finite in the z direction, see figure 5.1. We will use the same ansatz as in chapter 4, but in this case the secular equation leads to a transcendental equation, which we have to solve numerically. Using this method, we will find both surface and bulk states, although for thin films the distinction is not as clear, since all states have a non-zero wave function throughout the material. For a very thick insulator we expect to get the same result as in the semi-infinite case. But as the thickness gets smaller the overlap of the wave functions of the states on opposite surfaces induces a gap. This can be seen as bonding/anti-bonding combinations of the surface states on opposite surfaces.

To illustrate the method used in the general case, we will first analyze the Hamiltonian at the gamma point. Here the Hamiltonian gets block diagonal, and we can analyze each block separately.

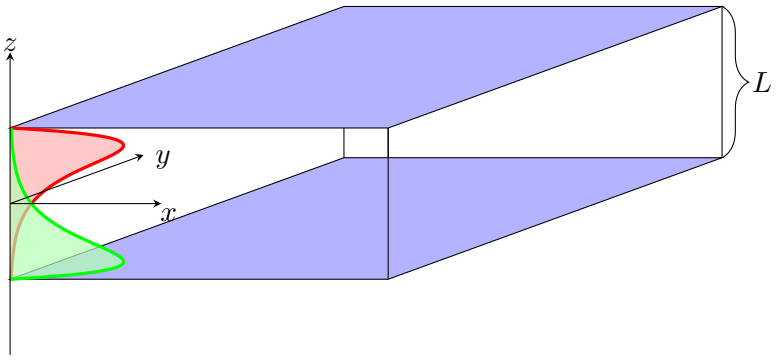


Figure 5.1: In a finite geometry, the overlap between wave function on either side of the insulator can induce a gap in the spectrum.

We consider a topological insulator, infinite in the x, y plane but with surfaces at $z = \pm\frac{L}{2}$. Again we will use hard-wall boundary conditions:

$$\Psi(z = \pm\tfrac{L}{2}) = 0 \tag{5.1}$$

5.1 Spectrum at the gamma point

We will start by finding the surface states at the gamma point, which will be useful when constructing an effective 2D model for the surface states. Furthermore, it illustrates the method used for $k_{||} \neq 0$. As in 4, we have the time independent Schrödinger equation:

$$H(k_x, k_y, -i\partial_z)\Psi(z) = E\Psi(z). \tag{5.2}$$

At the gamma point, the two spin blocks of the Hamiltonian decouples, and can be written as:

$$H_0(-i\partial_z) = H(k_x = 0, k_y = 0, -i\partial_z) = \begin{pmatrix} h & 0 \\ 0 & h^* \end{pmatrix}, \tag{5.3}$$

where the spin up block is given by:

$$h = \begin{pmatrix} C + M - D_-\partial_z^2 & -iA_1\partial_z \\ -iA_1\partial_z & C - M - D_+\partial_z^2 \end{pmatrix}. \tag{5.4}$$

We only need to solve one of the blocks, then the solutions to the other can be found by time reversal operation, or equivalently by replacing $A_1 \to -A_1$. Here we will solve the spin up block. Again we use the ansatz $\psi_\lambda e^{\lambda z}$, which gives the same solutions $\lambda = \beta\lambda_\alpha$, from eq. (4.15), only with $k_{||} = 0$. We only get one eigen-spinor for the spin up block:

$$\psi^\uparrow_{\alpha\beta}(z) = \begin{pmatrix} J_{\alpha+} \\ -iA_1\lambda_\alpha \end{pmatrix} e^{\beta\lambda_\alpha z} \tag{5.5}$$

where $J_{\alpha+} = D_+\lambda_\alpha^2 - C + M + E$. Since we have inversion symmetry in this case, it is convenient to change to basis states which are eigenstates of the inversion operator:

$$\varphi^\uparrow_{\alpha+}(z) = \frac{1}{2}(\psi^\uparrow_{\alpha+}(z) + \psi^\uparrow_{\alpha-}(z)) = \begin{pmatrix} J_{\alpha+}\cosh(\lambda_\alpha z) \\ -iA_1\lambda_\alpha \sinh(\lambda_\alpha z) \end{pmatrix} \tag{5.6}$$

$$\varphi^\uparrow_{\alpha-}(z) = \frac{1}{2}(\psi^\uparrow_{\alpha+}(z) - \psi^\uparrow_{\alpha-}(z)) = \begin{pmatrix} J_{\alpha+}\sinh(\lambda_\alpha z) \\ -iA_1\lambda_\alpha \cosh(\lambda_\alpha z) \end{pmatrix} \tag{5.7}$$

Here the second index denotes the inversion eigenvalue. Under inversion $\sinh(\lambda_\alpha z) \to -\sinh(\lambda_\alpha z)$, while $\cosh(\lambda_\alpha z) \to \cosh(\lambda_\alpha z)$. However, the inversion operator also operates in the 4-spinor space, given by $\sigma_0 \otimes \tau_z$, as we saw in section 3.3. Therefore, the above states, are actually inversion eigenstates.

A general solution to the spin up block can be written:

$$\Psi^\uparrow(z) = \sum_{\alpha,\xi} C^\uparrow_{\alpha\xi}\varphi^\uparrow_{\alpha\xi}(z). \tag{5.8}$$

Now we impose the boundary conditions:

$$\Psi\left(z = \pm\frac{L}{2}\right) = 0. \tag{5.9}$$

This gives a linear homogeneous system of equations for the coefficients:

$$\begin{pmatrix} J_{1+}\cosh(\frac{\lambda_1 L}{2}) & J_{2+}\cosh(\frac{\lambda_2 L}{2}) & J_{1+}\sinh(\frac{\lambda_1 L}{2}) & J_{2+}\sinh(\frac{\lambda_2 L}{2}) \\ -iA_1\lambda_1\sinh(\frac{\lambda_1 L}{2}) & -iA_1\lambda_2\sinh(\frac{\lambda_2 L}{2}) & -iA_1\lambda_1\cosh(\frac{\lambda_1 L}{2}) & -iA_1\lambda_2\cosh(\frac{\lambda_2 L}{2}) \\ J_{1+}\cosh(\frac{\lambda_1 L}{2}) & J_{2+}\cosh(\frac{\lambda_2 L}{2}) & -J_{1+}\sinh(\frac{\lambda_1 L}{2}) & -J_{2+}\sinh(\frac{\lambda_2 L}{2}) \\ iA_1\lambda_1\sinh(\frac{\lambda_1 L}{2}) & iA_1\lambda_2\sinh(\frac{\lambda_2 L}{2}) & -iA_1\lambda_1\cosh(\frac{\lambda_1 L}{2}) & -iA_1\lambda_2\cosh(\frac{\lambda_2 L}{2}) \end{pmatrix} \begin{pmatrix} C_{1+} \\ C_{2+} \\ C_{1-} \\ C_{2-} \end{pmatrix} = 0. \tag{5.10}$$

This equation can be reduced by simple row operations to

$$\begin{pmatrix} J_{1+}\cosh(\frac{\lambda_1 L}{2}) & J_{2+}\cosh(\frac{\lambda_2 L}{2}) & 0 & 0 \\ \lambda_1\sinh(\frac{\lambda_1 L}{2}) & \lambda_2\sinh(\frac{\lambda_2 L}{2}) & 0 & 0 \\ 0 & 0 & J_{1+}\sinh(\frac{\lambda_1 L}{2}) & J_{2+}\sinh(\frac{\lambda_2 L}{2}) \\ 0 & 0 & \lambda_1\cosh(\frac{\lambda_1 L}{2}) & \lambda_2\cosh(\frac{\lambda_2 L}{2}) \end{pmatrix} \begin{pmatrix} C_{1+} \\ C_{2+} \\ C_{1-} \\ C_{2-} \end{pmatrix} = 0. \tag{5.11}$$

This matrix is now block diagonal, with each block corresponding to one inversion eigenvalue. For the upper block, corresponding to positive inversion eigenvalue, we get the secular equation:

$$\begin{aligned} \frac{J_{1+}\lambda_2}{J_{2+}\lambda_1} &= \frac{\tanh(\frac{\lambda_1 L}{2})}{\tanh(\frac{\lambda_2 L}{2})} \\ \Leftrightarrow \frac{(D_+\lambda_1^2 - C + M + E)\lambda_2}{(D_+\lambda_2^2 - C + M + E)\lambda_1} &= \frac{\tanh(\frac{\lambda_1 L}{2})}{\tanh(\frac{\lambda_2 L}{2})}. \end{aligned} \tag{5.12}$$

Here, $\lambda_{1,2}$ are dependent on the energy only, and the solutions to this equation gives all energies of the $k_{||} = 0$ states, which are even under inversion. The other block is the same just with all cosh and sinh interchanged, and we get a similar secular equation for the states at the gamma point, which are odd under inversion:

$$\frac{(D_+\lambda_1^2 - C + M + E)\lambda_2}{(D_+\lambda_2^2 - C + M + E)\lambda_1} = \frac{\tanh(\frac{\lambda_2 L}{2})}{\tanh(\frac{\lambda_1 L}{2})}. \tag{5.13}$$

By solving eq. (5.12) and (5.13), we can obtain the spectrum of at the gamma point. These equations are transcendental and we will solve them numerically. Solving these equations is equivalent to finding the zero points of the functions:

$$g_{\text{even}}(E) = \frac{(D_+\lambda_2(E)^2 - C + M + E)\tanh(\frac{\lambda_1(E)L}{2})}{\lambda_2(E)} - \frac{(D_+\lambda_1(E)^2 - C + M + E)\tanh(\frac{\lambda_2(E)L}{2})}{\lambda_1(E)}. \tag{5.14}$$

$$g_{\text{odd}}(E) = \frac{(D_+\lambda_2(E)^2 - C + M + E)\tanh(\frac{\lambda_2(E)L}{2})}{\lambda_2(E)} - \frac{(D_+\lambda_1(E)^2 - C + M + E)\tanh(\frac{\lambda_1(E)L}{2})}{\lambda_1(E)}. \tag{5.15}$$

Since λ_1 and λ_2 are in general complex numbers, these functions are complex-valued. To simplify our problem, we will instead take the sum of the real and imaginary parts of these

functions, and then find the zeroes of the resulting real valued functions:

$$f_{\text{even}}(E) = \text{Re}[g_{\text{even}}(E)] + \text{Im}[g_{\text{even}}(E)] \tag{5.16}$$

$$f_{\text{odd}}(E) = \text{Re}[g_{\text{odd}}(E)] + \text{Im}[g_{\text{odd}}(E)] \tag{5.17}$$

We find the zeroes by iterating over an array of energies, and checking whether the product of to subsequent values of the function is negative. If it is, then either the function has a zero or a divergency between the two points. By setting a threshold value for the difference between the value of the function at these points we avoid the divergencies. At the zeroes of these functions, we know that a non-trivial solution for the coefficients exists, which gives a non-zero wave function, if the basis functions are linearly independent. The only possibility for a solution to the secular equations where the basis states are linearly dependent is if $\lambda_1 = \lambda_2$, $\lambda_1 = 0$ or $\lambda_2 = 0$. The case $\lambda_1 = \lambda_2$ happens when R as defined in eq. (4.17) is zero. Therefore, if we find a zero of one of the above functions, we check whether R also has a zero. If $F^2 - R = 0$ then either λ_1 or λ_2 is zero, which can be seen by calculating $\lambda_1\lambda_2 = \frac{\sqrt{F^2-R}}{2|D_+D_-|}$ from eq. (4.15). Therefore, we also check whether the function $F^2 - R$ has a zero. The energies where f_{even} or f_{odd} are zero, but both R and $F^2 - R$ are non-zero, are the eigenenergies of the finite system. See figure 5.2 for a plot of these functions. This way we can iterate over a

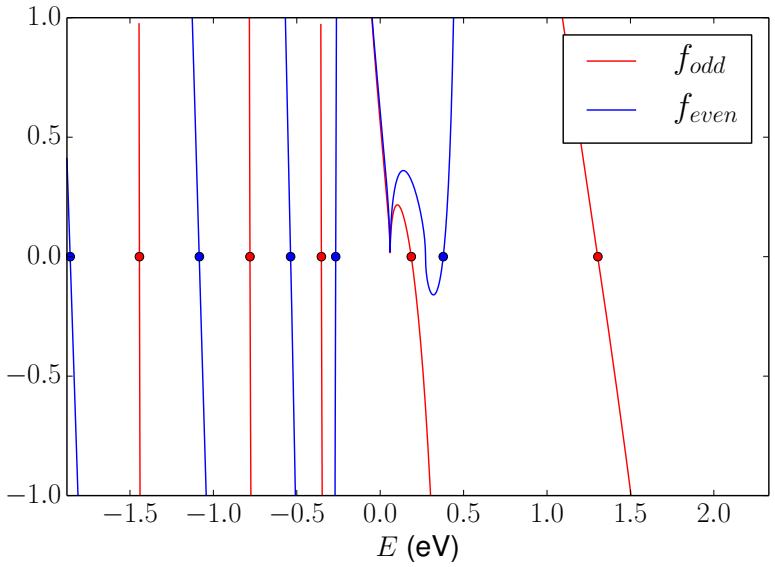


Figure 5.2: The zeroes of the functions $f_{\text{even}}(E)$ and f_{odd} correspond to either eigenenergies of the system or points where our basis states are linearly dependent. To get only the eigenenergies, we drop all the zeroes of $f_{\text{even/odd}}$ where either R or $F^2 - R$ is zero as well. The circles indicates the eigenenergies, found by this method. This example is for a thickness of $L = 20\,\text{Å}$ using the parameters from [11].

range of thicknesses and see how the gap depends on the thickness. Here we assume that the minimal gap is at the gamma point, which we will see when we calculate the full spectrum. Since surface states decays exponentially, and the gap is induced by the overlap of surface states at opposite surfaces, we expect an exponential decay of the gap as a function of the thickness. It turns out that for certain model parameters, there is an oscillation of the gap on

top of this exponential decay, as seen on figure 5.3a. This oscillation is also discussed in [23], [24] and [25].

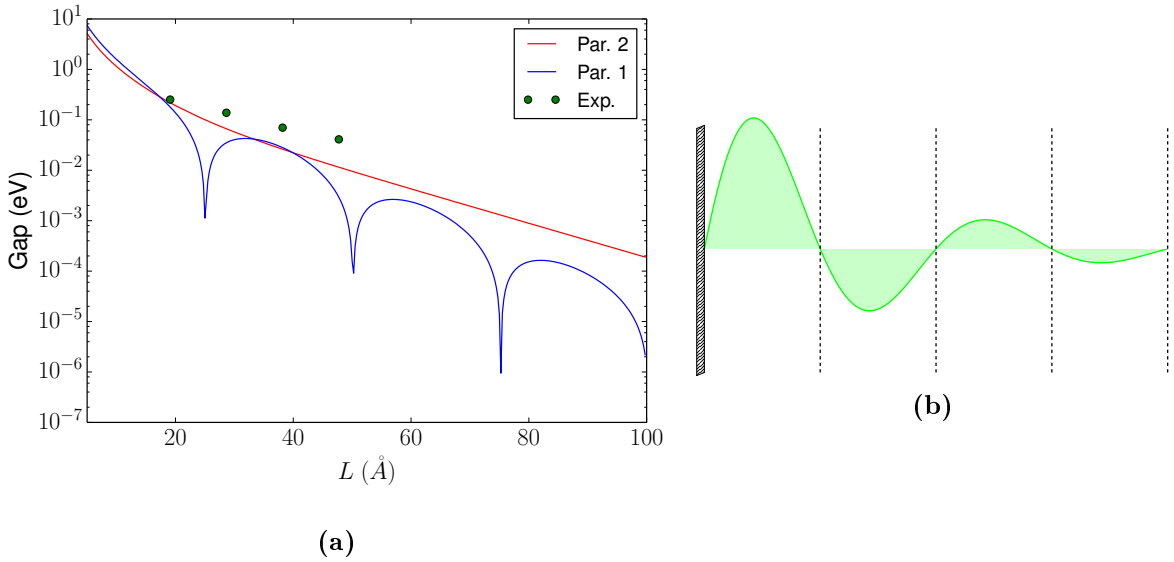


Figure 5.3: (a) Gap between the surface states as a function of the thickness. The red line is calculated using the model parameters from [11], and shows only an exponential decay, while the blue line is calculated using the parameters from [12] shows an oscillation on top of the exponential decay. We emphasize that the gap actually goes to zero, at local minima of the blue curve, which is not seen, because of the finite resolution. The zeroes of the gap occurs at integer multiples of 25 Å. The green circles are experimental data from [26]. (b) Oscillating wave function for a semi-infinite topological insulator. We can introduce another surface, exactly at one of the zeroes of the wavefunction, indicated by the dashed lines. Then, this wave function, will still be fulfill the new boundary condition. Hence, the energy is not changed, by introducing a surface at one of these points.

The equidistant zeroes of the gap, when using the model parameters from [12] can be explained by looking at the wave functions in the semi infinite case. For these model parameters the wave function are oscillating, which means that the wave function is zero at equidistant points within the insulator. If we make a second surface exactly at one of these points, then the wave function of the semi-infinite case is still a solution to the finite insulator. This situation is shown qualitatively in figure 5.3b. Therefore, since the semi-infinite insulator is gapless, the finite insulator must be gapless too. Furthermore, we note here that each time the gap goes to zero the energies of the even and odd surface state interchanges, such that if we define the gap by $\Delta = E_- - E_+$, it changes sign.

In [26], the finite size effect were investigated by ARPES experiments on films of Bi_2Se_3, with thicknesses of 2-6 quintuple layers (9.547 Å). The measured gap is plotted in figure 5.3a, for 2-5 quintuple layers. For 6 quintuple layers they do not see any gap in the experiment. The experiment clearly shows an exponential decay of the gap, as a function of thickness. The size of the gap, agrees remarkably well at two quintuple layers, with the theoretical predictons for both sets of parameters, but decays slower with the thickness. The experimental data show no sign of oscillation of the gap. This strongly suggest using the parameters from [11]. However, absence of gap oscillation is not conclusive, since we only have 4 points. It is an experimental challenge to fabricate thin films, with a non-integer number of quintuple layers, due to the strong coupling within the quintuple layers.

5.1.1 Wave functions at the gamma point

Now we will find the wave functions at the gamma point. To do this we have to solve eq. (5.11) for the coefficients. Since each block is only 2×2 this can be done analytically, however we still need to find the energy numerically. For the upper block corresponding to positive inversion eigenvalue, we can choose the coefficients to be:

$$C_{1+} = -\frac{1}{\lambda_1 \sinh(\frac{\lambda_1 L}{2})}, \quad C_{2+} = \frac{1}{\lambda_2 \sinh(\frac{\lambda_2 L}{2})}. \tag{5.18}$$

Which gives the eigenstate (written as a 2-spinor in the spin up subspace):

$$\begin{aligned}
\varphi_+(z) &= N_+ \left(-\frac{1}{\lambda_1 \sinh(\frac{\lambda_1 L}{2})} \varphi^{\uparrow}_{1+}(z) + \frac{1}{\lambda_2 \sinh(\frac{\lambda_2 L}{2})} \varphi^{\uparrow}_{1+}(z) \right) \\
&= N_+ \begin{pmatrix} -J_{1+} \frac{\cosh(\lambda_1 z)}{\lambda_1 \sinh(\frac{\lambda_1 L}{2})} + J_{2+} \frac{\cosh(\lambda_2 z)}{\lambda_2 \sinh(\frac{\lambda_2 L}{2})} \\ iA_1 \left(\frac{\sinh(\lambda_1 z)}{\sinh(\frac{\lambda_1 L}{2})} - \frac{\sinh(\lambda_2 z)}{\sinh(\frac{\lambda_2 L}{2})} \right) \end{pmatrix} \\
&= N_+ \begin{pmatrix} -\frac{J_{1+}}{\tanh(\frac{\lambda_1 L}{2})} \left(\frac{\cosh(\lambda_1 z)}{\cosh(\frac{\lambda_1 L}{2})} - \frac{\cosh(\lambda_2 z)}{\cosh(\frac{\lambda_2 L}{2})} \right) \\ iA_1 \left(\frac{\sinh(\lambda_1 z)}{\sinh(\frac{\lambda_1 L}{2})} - \frac{\sinh(\lambda_2 z)}{\sinh(\frac{\lambda_2 L}{2})} \right) \end{pmatrix} \\
&= N_+ \begin{pmatrix} D_+ \left(\frac{\lambda_2^2 - \lambda_1^2}{\lambda_1 \tanh(\frac{\lambda_1 L}{2}) - \lambda_2 \tanh(\frac{\lambda_2 L}{2})} \right) \left(\frac{\cosh(\lambda_1 z)}{\cosh(\frac{\lambda_1 L}{2})} - \frac{\cosh(\lambda_2 z)}{\cosh(\frac{\lambda_2 L}{2})} \right) \\ iA_1 \left(\frac{\sinh(\lambda_1 z)}{\sinh(\frac{\lambda_1 L}{2})} - \frac{\sinh(\lambda_2 z)}{\sinh(\frac{\lambda_2 L}{2})} \right) \end{pmatrix}
\end{aligned} \tag{5.19}$$

To rewrite the wave function into this simple form we used eq. (5.12), in both steps. We have introduced a normalization constant $N_+^{\uparrow}$. To find λ_1 and λ_2 we have to find the energies by solving the corresponding secular equation. In eq. (5.11) we see that the odd inversion block is equal to the even inversion block if we interchange sinh and cosh. So the wave functions of the odd inversion eigenstates can be obtained by simply interchanging sinh and cosh in the even inversion wave function:

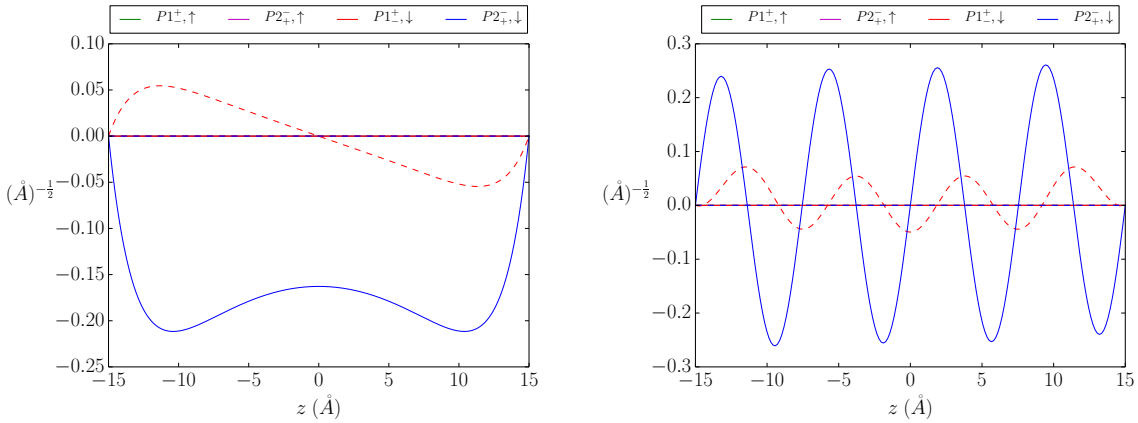
$$\varphi_-(z) = N_- \begin{pmatrix} D_+ \left(\frac{\lambda_2^2 - \lambda_1^2}{\lambda_1 \coth(\frac{\lambda_1 L}{2}) - \lambda_2 \coth(\frac{\lambda_2 L}{2})} \right) \left(\frac{\sinh(\lambda_1 z)}{\sinh(\frac{\lambda_1 L}{2})} - \frac{\sinh(\lambda_2 z)}{\sinh(\frac{\lambda_2 L}{2})} \right) \\ iA_1 \left(\frac{\cosh(\lambda_1 z)}{\cosh(\frac{\lambda_1 L}{2})} - \frac{\cosh(\lambda_2 z)}{\cosh(\frac{\lambda_2 L}{2})} \right) \end{pmatrix} \tag{5.20}$$

The two degenerate solutions are written as full 4-spinors given by:

$$\Psi^{\uparrow}_{\xi}(z) = \begin{pmatrix} \varphi_{\xi}(z) \\ 0 \end{pmatrix} \text{ and } \Psi^{\downarrow}_{\xi}(z) = \begin{pmatrix} 0 \\ \tau_z \varphi_{\xi}(z) \end{pmatrix}, \tag{5.21}$$

where the energy is found by solving the secular equation corresponding to the inversion eigenvalue ξ. For convenience we have written the spin down state with a τ_z, instead of complex conjugating[1]. Some examples of these wave functions are seen on figure 5.4. We see

[1] The i in the second component, becomes $-i$ while complex conjugating. The rest of the spinor is either real, if λ_1 and λ_2 are both purely real or imaginary, or imaginary if $\lambda_1 = \lambda_2^*$ is complex. To avoid getting an



(a) The wave function of the lower surface state, having an energy of 0.21 eV.

(b) Wave function of a bulk state at the energy −0.70 eV

Figure 5.4: The wave functions in the spin down block, for a thickness of $L = 30$ Å. The color denotes the 4 different components, in the basis of our model Hamiltonian. The color denotes the 4 different components, in the basis of our model Hamiltonian. The solid lines are the real part, while dashed are the imaginary part.

that the components of the bulk state looks like the wave functions of an infinite square well, where the number of oscillations increases with energy. In contrast the surface states have large peaks close to both surfaces, which becomes more significant as L is increased. On figure 5.5, the density of the surface state with the highest energy, is shown for a range of thicknesses L. Here we see that the density at $z = 0$ decreases, when L is increased, and eventually becomes negligible. This agrees with the fact that surface states on opposite surfaces decouples in the large L limit, and the gap becomes negligible. When L is large enough that the gap can be neglected, then all four surface states are degenerate at the gamma point. Then we can make a superpositions of the even and odd states, which will be localized close to a single surface, simmilar to the wave functions found in chapter 4.

5.1.2 2D model

Using the states at the gamma point as a basis, we can compute an effective 2D model of a film of 3D topological insulator. We follow the approach of [16]. We will only use the four surface states, denoted by $|\Psi^\sigma_\xi\rangle$, where ξ denotes the inversion eigenvalue, and σ the spin. We denote the energies of these states E_ξ, where ξ is the inversion eigenvalue again. We split the Hamiltonian into the $k_{||} = 0$ part and a part dependent on k_x, k_y:

$$H = H_0(-i\partial_z) + \Delta H(k_x, k_y) \tag{5.22}$$

The momentum dependent part can be written:

$$\Delta H(k_x, k_y) = D_2 k_{||}^2 - B_2 k_{||}^2 \sigma_0 \otimes \tau_z + A_2(k_x\sigma_x + k_y\sigma_y) \otimes \tau_x \tag{5.23}$$

overall sign in the latter case, we use τ_z instead of complex conjugation.

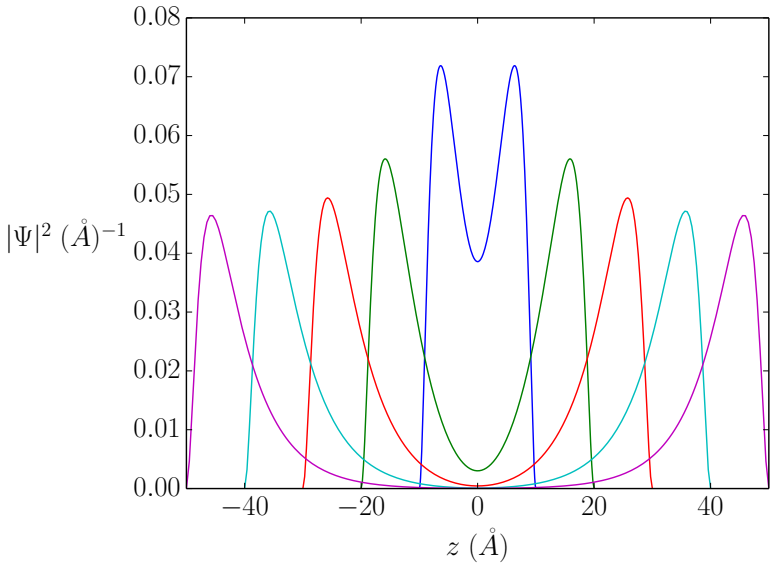


Figure 5.5: Density of the upper surface state for different thicknesses from $L = 10\,\text{Å}$ to $L = 100\,\text{Å}$. For small L we see a big change in the density, when changing L, while we can the density for $L = 100\,\text{Å}$ almost has the same form the density for $L = 80\,\text{Å}$.

Since our basis is already eigenstates of H_0 we need only to calculate the matrix elements of ΔH. By using the inversion symmetry of the Hamiltonian, $\Delta H(k_x, k_y) = I\Delta H(-k_x, -k_y)I$, and the fact that our basis states are eigenstates of the inversion operator we get:

$$\langle \Psi^{\sigma'}_{\xi'} | \Delta H(k_x, k_y) | \Psi^{\sigma}_{\xi} \rangle = \langle \Psi^{\sigma'}_{\xi'} | I\Delta H(-k_x, -k_y) I | \Psi^{\sigma}_{\xi} \rangle = \xi\xi' \langle \Psi^{\sigma'}_{\xi'} | \Delta H(-k_x, -k_y) | \Psi^{\sigma}_{\xi} \rangle \tag{5.24}$$

$$\tag{5.25}$$

From this we conclude that the $k_{||}^2$ terms only couple states of the same parity, while the terms linear in momentum couple states of opposite parity. The only non-zero matrix elements are:

$$\langle \Psi^{\sigma}_{\xi} | \Delta H(k_x, k_y) | \Psi^{\sigma}_{\xi} \rangle = D_2 k_{||}^2 - B_2 k_{||}^2 \langle \varphi_\xi | \tau_z | \varphi_\xi \rangle \tag{5.26}$$

$$\langle \Psi^{\uparrow}_{+} | \Delta H(k_x, k_y) | \Psi^{\downarrow}_{-} \rangle = A_2 k_- \langle \varphi_+ | \tau_x \tau_z | \varphi_- \rangle \tag{5.27}$$

$$\langle \Psi^{\uparrow}_{-} | \Delta H(k_x, k_y) | \Psi^{\downarrow}_{+} \rangle = A_2 k_- \langle \varphi_- | \tau_x \tau_z | \varphi_+ \rangle \tag{5.28}$$

By defining some new quantities:

$$B = B_2 \frac{\langle \varphi_- | \tau_z | \varphi_- \rangle - \langle \varphi_+ | \tau_z | \varphi_+ \rangle}{2} \tag{5.29}$$

$$D = B_2 \frac{\langle \varphi_- | \tau_z | \varphi_- \rangle + \langle \varphi_+ | \tau_z | \varphi_+ \rangle}{2} - D_2 \tag{5.30}$$

$$E_0 = \frac{E_+ - E_-}{2} \tag{5.31}$$

$$\Delta = E_- - E_+ \tag{5.32}$$

$$\tilde{A}_2 = A_2 \langle \varphi_- | \tau_x \tau_z | \varphi_+ \rangle = -iA_2 \langle \varphi_- | \tau_y | \varphi_+ \rangle \tag{5.33}$$

We can write the effective model in the basis $\{|\Psi_-^\uparrow\rangle, |\Psi_+^\downarrow\rangle, |\Psi_+^\uparrow\rangle, |\Psi_-^\downarrow\rangle\}$:

$$H_{2D}(k_x, k_y) = E_0 - Dk_{||}^2 + \begin{pmatrix} \frac{\Delta}{2} - Bk_{||}^2 & \tilde{A}_2 k_- & 0 & 0 \\ \tilde{A}_2^* k_+ & -\frac{\Delta}{2} + Bk_{||}^2 & 0 & 0 \\ 0 & 0 & -\frac{\Delta}{2} + Bk_{||}^2 & -\tilde{A}_2^* k_- \\ 0 & 0 & -\tilde{A}_2 k_+ & \frac{\Delta}{2} - Bk_{||}^2 \end{pmatrix} \tag{5.34}$$

In [16], it is argued, that the parameter $\tilde{A}_2$, is either purely real or purely imiginary, giving two different cases for this model. However, the basis states we use have an arbitrary phase or gauge freedom. Hence, by performing a gauge transformation only on the positive inversion states, $|\varphi_+\rangle \to e^{i\theta}|\varphi_+\rangle$, we get the parameter $\tilde{A}_2 \to e^{i\theta}\tilde{A}_2$. We see that the two cases are related by this simple transformation, which does not change any of the other parameters, since they include only matrix elements between states with the same inversion eigenvalue.

This model is equivalent to the BHZ model for HgTe quantum wells, derived in [5]. This model was shown to describe a 2D topological insulator. Using the same approach, that we used in chapter 4 to find surface states, one can find edge states in the BHZ model. Therefore, a thin film of a 3D topological insulator, can behave as 2D topological insulator. Here the surface states are gapped out, because of the coupling between the two surfaces, and edge states arises within the gap. Whether the thin film is a 2D topological insulator or not depends on the parameters of the model. It can be shown that for;

$$\frac{\Delta}{2B} > 0, \tag{5.35}$$

the thin film is topologically non-trivial, and there exists counterpropagating edge states. In appendix A, we show the equivalence of these models, calculate the edge states and derive the condition in eq. 5.35 for the existence of edge states. For the parameters of [12], the gap Δ oscillates, as a function of the thickness L. Therefore, the thin film oscillates between the trivial and non-trivial topology, assuming that the sign of B does not change, which was showed in [23].

5.2 Bulk and surface spectrum

Now we will calculate the full spectrum of a finite topological insulator. We will do this, and the rest of this chapter, only using the parameters from [11].

We will change to a basis of states where the state at in plane momentum k_x, k_y is related by inversion to the same state at $-k_x, -k_y$. In chapter 4 we used the basis:

$$\psi_{\alpha\beta 1} e^{i\beta\lambda_\alpha z} = \begin{pmatrix} J_{\alpha+} \\ -iA_1\beta\lambda_\alpha \\ 0 \\ A_2 k_+ \end{pmatrix} e^{i\beta\lambda_\alpha z}, \quad \psi_{\alpha\beta 2} e^{i\beta\lambda_\alpha z} = \begin{pmatrix} A_2 k_- \\ 0 \\ i\beta A_1 \lambda \\ J_{\alpha-} \end{pmatrix} e^{i\beta\lambda_\alpha z}, \tag{5.36}$$

where $\beta = \pm$ and $\alpha = 1, 2$. Now we change the basis to $\varphi_{\alpha\xi\gamma} = \frac{1}{2}(\psi_{\alpha+\gamma} + (-1)^{\gamma-1}(\xi\psi_{\alpha-\gamma}))$.

The positive inversion basis states are given by:

$$\varphi_{\alpha+1}(z) = \begin{pmatrix} J_{\alpha+}\cosh(\lambda_\alpha z) \\ -iA_1\lambda_\alpha \sinh(\lambda_\alpha z) \\ 0 \\ A_2 k_+ \cosh(\lambda_\alpha z) \end{pmatrix}, \quad \varphi_{\alpha+2}(z) = \begin{pmatrix} A_2 k_- \sinh(\lambda_\alpha z) \\ 0 \\ iA_1\lambda \cosh(\lambda_\alpha z) \\ J_{\alpha-}.\sinh(\lambda_\alpha z) \end{pmatrix} \tag{5.37}$$

The negative inversion states have the same form except, all cosh are replaced by sinh and vice versa. The general solution to the Schrödinger equation is given by a superposition of these eight eigenstates:

$$\Psi(z) = \sum_{\alpha\xi\gamma} C_{\alpha\xi\gamma}\varphi_{\alpha\xi\gamma}(z) \tag{5.38}$$

When we impose the boundary conditions (5.1) it gives a 8 dimensional linear homogeneous system of equations for the coefficients. It will be convenient to arrange the coefficients in the order $\mathbf{C} = (C_{1+1}, C_{1-2}, C_{2+1}, C_{2-2}, C_{1-1}, C_{1+2}, C_{2-1}, C_{2+2})$. The system of equations can be written as:

$$M\mathbf{C} = 0 \tag{5.39}$$

M is a matrix where each column consist of the corresponding 4-spinor wavefunction evaluated at the boundaries $\varphi_{\alpha\beta\gamma}(\frac{L}{2}), \varphi_{\alpha\beta\gamma}(-\frac{L}{2})$. This matrix can be written simpler on block form:

$$M = \begin{pmatrix} m_1(\frac{L}{2}) & m_2(\frac{L}{2}) \\ m_1(-\frac{L}{2}) & m_2(-\frac{L}{2}) \end{pmatrix}, \tag{5.40}$$

where

$$m_1(\tfrac{L}{2}) = \begin{pmatrix} J_{1+}\cosh(\frac{\lambda_1 L}{2}) & A_2 k_- \cosh(\frac{\lambda_1 L}{2}) & J_{2+}\cosh(\frac{\lambda_2 L}{2}) & A_2 k_- \cosh(\frac{\lambda_2 L}{2}) \\ -iA_1\lambda_1 \sinh(\frac{\lambda_1 L}{2}) & 0 & -iA_1\lambda_2 \sinh(\frac{\lambda_2 L}{2}) & 0 \\ 0 & iA_1\lambda_1 \sinh(\frac{\lambda_1 L}{2}) & 0 & iA_1\lambda_2 \sinh(\frac{\lambda_2 L}{2}) \\ A_2 k_+ \cosh(\frac{\lambda_1 L}{2}) & J_{1-}\cosh(\frac{\lambda_1 L}{2}) & A_2 k_+ \cosh(\frac{\lambda_2 L}{2}) & J_{2-}\cosh(\frac{\lambda_2 L}{2}) \end{pmatrix}. \tag{5.41}$$

and $m_2(\frac{L}{2})$ is obtained from $m_1(\frac{L}{2})$ by interchanging cosh and sinh.

Now we want to compute the determinant to obtain the secular equation for the nontrivial solution of the coefficients. If we just try and calculate the determinant we would get 4608 nonzero terms. But we can perform row operations to obtain a simpler matrix equation, equivalent to the original one. Using that sinh is odd and cosh is even, and multiplying row 5 and 6 by -1, we get the matrix:

$$\begin{pmatrix} m_1(\frac{L}{2}) & m_2(\frac{L}{2}) \\ m_1(\frac{L}{2}) & -m_2(\frac{L}{2}) \end{pmatrix} \mathbf{C} = 0 \tag{5.42}$$

This matrix can easily be reduced to a block diagonal matrix of the form:

$$\begin{pmatrix} m_1(\frac{L}{2}) & 0 \\ 0 & m_2(\frac{L}{2}) \end{pmatrix} \mathbf{C} = 0 \tag{5.43}$$

The secular equation for nontrivial solutions now factors into the product of the determinants of the two submatrices. In addition, since this matrix is on block form, we can find solutions, using only four of the basis states. First we compute the determinant of the upper block:

$$\begin{aligned}
&\det(m_1(\tfrac{L}{2})) = \\
&\begin{vmatrix} J_{1+}\cosh(\lambda_1 L/2) & A_2 k_- \cosh(\lambda_1 L/2) & J_{2+}\cosh(\lambda_2 L/2) & A_2 k_- \cosh(\lambda_2 L/2) \\ -iA_1\lambda_1 \sinh(\lambda_1 L/2) & 0 & -iA_1\lambda_2 \sinh(\lambda_2 L/2) & 0 \\ 0 & iA_1\lambda_1 \sinh(\lambda_1 L/2) & 0 & iA_1\lambda_2 \sinh(\lambda_2 L/2) \\ A_2 k_+ \cosh(\lambda_1 L/2) & J_{1-}\cosh(\lambda_1 L/2) & A_2 k_+ \cosh(\lambda_2 L/2) & J_{2-}\cosh(\lambda_2 L/2) \end{vmatrix} \\
&= A_1^2 \sinh(\lambda_1 L/2)\sinh(\lambda_2 L/2)\cosh(\lambda_1 L/2)\cosh(\lambda_2 L/2)\lambda_1\lambda_2(2A_2^2k^2 - J_{2+}J_{1-} - J_{1+}J_{2-}) \\
&+ A_1^2 \sinh^2(\lambda_1 L/2)\cosh^2(\lambda_2 L/2)\lambda_1^2(J_{2+}J_{2-} - A_2^2k^2) \\
&+ A_1^2 \sinh^2(\lambda_2 L/2)\cosh^2(\lambda_1 L/2)\lambda_2^2(J_{1+}J_{1-} - A_2^2k^2) \\
&= A_1^2\lambda_1\lambda_2 \sinh(\lambda_1 L/2)\sinh(\lambda_2 L/2)\cosh(\lambda_1 L/2)\cosh(\lambda_2 L/2)\big(D_+D_-(\lambda_1^2-\lambda_2^2)^2 + A_1^2(\lambda_1^2+\lambda_2^2)\big) \\
&- A_1^4\lambda_1^2\lambda_2^2\big(\sinh^2(\lambda_1 L/2)\cosh^2(\lambda_2 L/2) + \sinh^2(\lambda_2 L/2)\cosh^2(\lambda_1 L/2)\big)
\end{aligned}$$

setting this equal to zero gives the equation:

$$\frac{\frac{D_+D_-}{A_1^2}(\lambda_1^2-\lambda_2^2)^2 + (\lambda_1^2+\lambda_2^2)}{\lambda_1\lambda_2} = \frac{\tanh(\lambda_1 L/2)}{\tanh(\lambda_2 L/2)} + \frac{\tanh(\lambda_2 L/2)}{\tanh(\lambda_1 L/2)}, \tag{5.44}$$

which was also found in [16]. This is the secular equation for solutions belonging to the upper block. Remember that $\lambda_{1,2}$ are functions of the in plane momentum $k_{||}$ and the energy E. A solution $E(k_{||})$ to this equation means that there is an eigenstate of the system at that point using only the first four basis states. This state can be either a surface state or a bulk state. Since the lower block $m_2(\frac{L}{2})$ can be obtained from the upper block, by interchanging cosh and sinh, we can simply do that substitution in the resulting secular equation, eq. (5.44). But this substitution does not change the equation, and the secular equation for solutions belonging to each block is the same. This fact ensures that all bands are at doubly degenerate, since any solution $E(k_{||})$ to eq. (5.44) ensures that there exist a nontrivial solution for the coefficients belonging to each block separately, with energy E and momentum $k_{||}$. This is actually what we expect, since we have a system with both inversion and time reversal symmetry. Therefore all bands must have a double degeneracy as explained in 2.9.2. And to find the wave function, it is convenient to just do it for one block, and then apply the TI symmetry operator to get state of the other block. This ensures that the two degenerate states are orthogonal, even though we use a basis which is not orthogonal.

Now we want to solve eq. (5.44) numerically. Hence, the problem now is finding the zeroes

in the $(k_{||}, E)$ plane of the function:

$$f(k_{||}, E) = \frac{\frac{D_+ D_-}{A_1^2}\left[\lambda_1^2(k_{||}, E) - \lambda_2^2(k_{||}, E)\right]^2 + (\lambda_1^2(k_{||}, E) + \lambda_2^2(k_{||}, E))}{\lambda_1(k_{||}, E)\lambda_2(k_{||}, E)} \tag{5.45}$$

$$-\frac{\tanh(\lambda_1(k_{||}, E)L/2)}{\tanh(\lambda_2(k_{||}, E)L/2)} - \frac{\tanh(\lambda_2(k_{||}, E)L/2)}{\tanh(\lambda_1(k_{||}, E)L/2)} \tag{5.46}$$

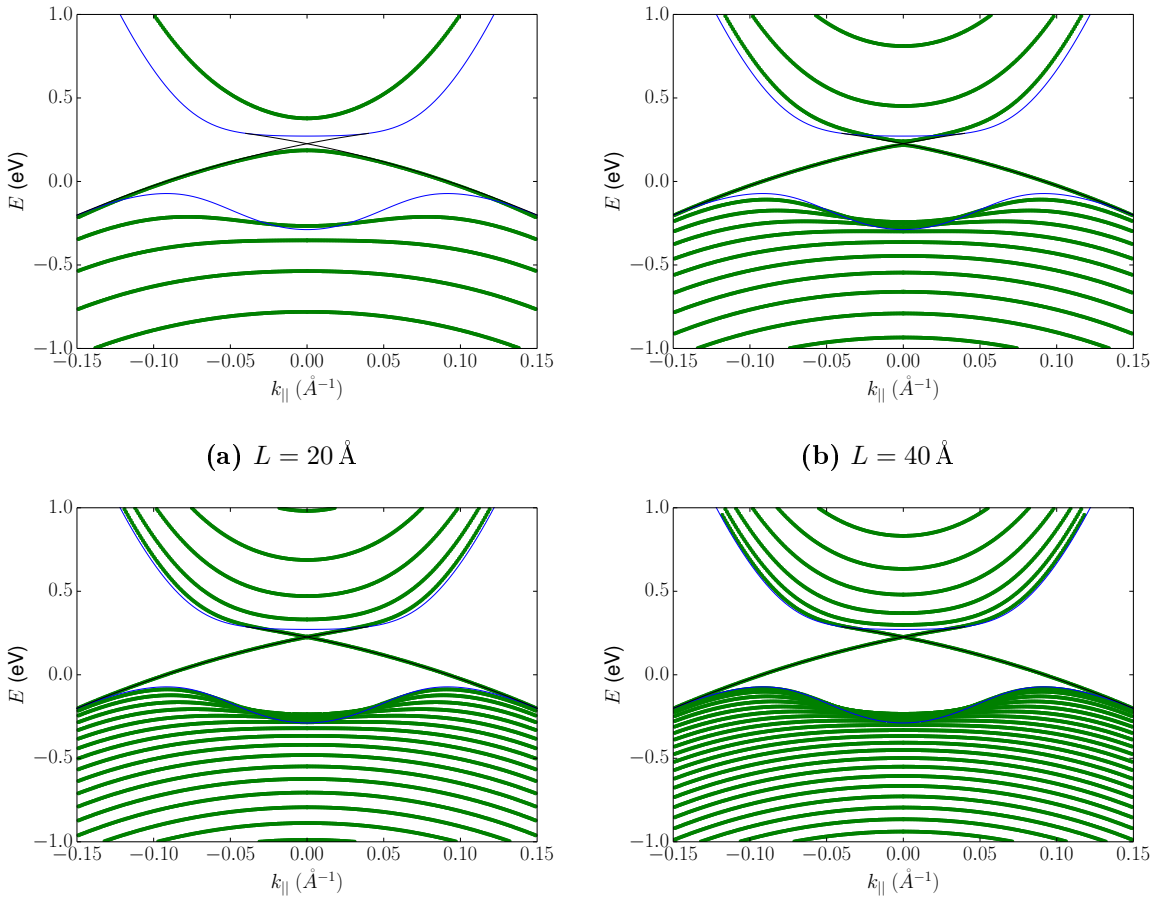
To do this, we iterate over $k_{||}$ and for each value k_i we find the points where $f(k_i, E)$ changes sign as a function of E. But since the function diverges for some points it can also change sign at those points, we impose a maximum threshold difference between the points on either side of the sign change. The necessary resolution and threshold were determined simply by looking at the plot of $f(k_i, E)$ for some specific k_i. This ad hoc method was necessary since we do not beforehand know, how many solutions to expect. By this method, we can obtain the spectrum. This method should give us all the points, where we have a non-trivial eigenstate, fulfilling the boundary conditions, but only if all our basis states are linearly independent. If we have some linearly dependent basis states, then we can make a linear combination, which is equal to zero, even though the coefficients are nonzero, by definition of linear dependence. The basis vectors are linearly dependent only if $A_1\lambda_1 = 0$, $A_1\lambda_2 = 0$ or if $\lambda_1 = \lambda_2$. We will again assume $A_1 \neq 0$, and the points where λ_1 or λ_2 are zero are avoided, since we divide by $\lambda_1\lambda_2$ in eq. (5.46). If $\lambda_1 = \lambda_2$ then by eq. (4.15), then $R = 0$. If we at each point where f changes sign check whether R changes sign and exclude the points where it does, then we avoid the solutions to the secular equation, which are caused by our basis becoming incomplete.

The resulting spectra are plotted in figure 5.6. We see that already at 60 Å the gap has closed and the surface bands coincide with the bands from the semi-infinite case. We also see that the spacing between the bulk bands decreases with the thickness, which is analogous to a 1D particle in a box, then the momentum gets quantized to fit the boundary condition.

5.2.1 Comparison with experimental data

Again we will compare to the results from the ARPES experiments, reported in [26]. The spectra for 1-6 quintuple layers are shown in figure 5.7. As indicated by blue and red dashed lines there doubly degenerate Dirac cone at the gamma point is split into two Dirac cones, slightly displaced from the gammapoint. This is due to a structural inversion asymmetry (SIA). The thin film is fabricated on a substrate, which breaks the inversion symmetry. One surface is subjected to vacuum, while the other surface is on top of the substrate. This effect is not included in our model, and we do not see this splitting. The dashed lines are the spectrum of the 2D model, including SIA. The outer branches of the upper surface states and the inner branches in the lower, are almost invisible in the data. The reason is that they are located primarily on the lower surface, and therefore are not reached by the incoming photons.

Besides this splitting due to SIA, the experimentally found spectrum, are qualitatively in great agreement with our calculations, shown in figure 5.6. We clearly see the particle-hole asymmetry; The upper surface band has a greater curvature than the lower, in both theory and experiment. Because of the smaller energy range in figure 5.7, we do not see the bulk bands until 5 and 6 quintuple layers. For 6 quintuple layers, we see two bulk bands above the



(a) $L = 20\,\text{Å}$ **(b)** $L = 40\,\text{Å}$

(c) $L = 60\,\text{Å}$ **(d)** $L = 80\,\text{Å}$

Figure 5.6: The dispersion relation for various thicknesses, calculated by numerically solving the secular equation. The blue lines shows the $k_z = 0$ bulk spectrum and the black is the surface spectrum for the semi infinite case for comparison. We see that the gap closes as the L increases, as well as the spacing between the bulk bands decreases.

Dirac point, with a spacing of approximately 0.1 eV at the gamma point, agreeing well with the spacing of 0.14 eV seen in figure 5.6c.

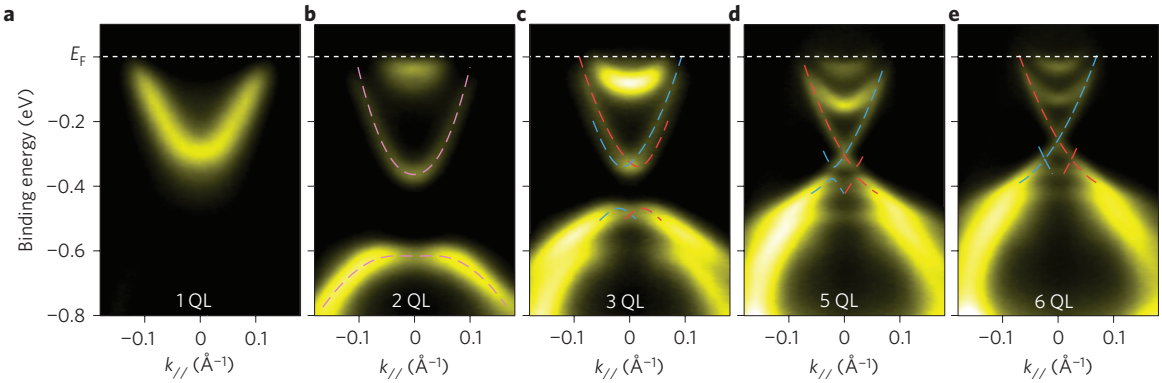


Figure 5.7: The spectra for thin films of Bi_2Se_3 measured by ARPES in [26], for different numbers of quintuple layers (QL). The dashed lines are the spectra of the 2D model, fitted to the data, including structural inversion asymmetry. Note that some of the brances are nearly invisible, which is a result of these states being located on the opposite surface.

5.3 Wave functions

To find the wave functions, we need to solve the system of equations for the coefficients, eq. (5.43). But since we already reduced it to block form we can simply solve one block at a time. We can also set $k_y = 0$ and then afterwards apply the rotation operator around the z axis to get to any general in-plane momentum vector k_x, k_y. This can be done, since the Hamiltonian to second order in momentum is invariant under any rotation around the z-axis. The equation we need to solve to find the coefficients is:

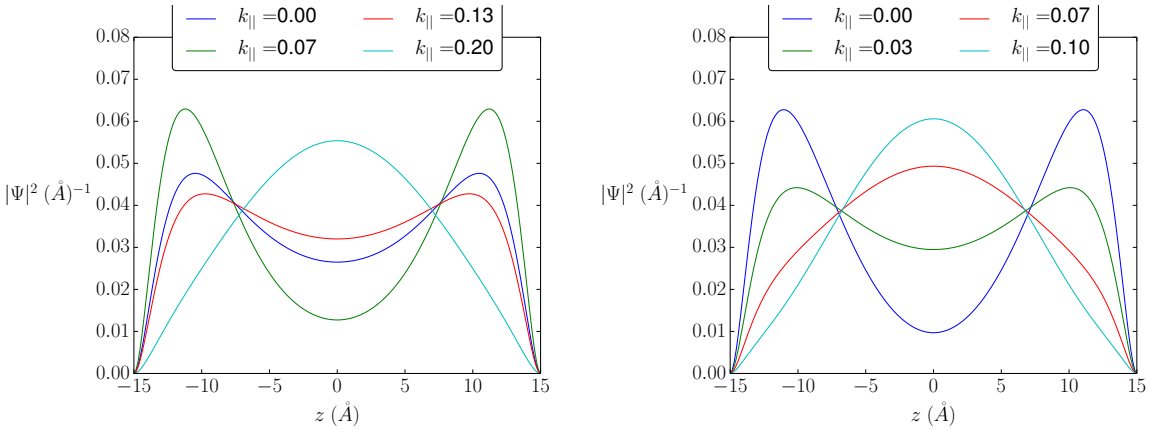
$$m_1(\tfrac{L}{2})\mathbf{C} = 0 \tag{5.47}$$

We know that this equation only has non-trivial solutions at the $k_{||}, E$ points found above. At these points we numerically do a *singular value decomposition* of the matrix. A singular value decomposition is a factorization of a matrix, similar to an eigenvalue decomposition, but can be done for any matrix. Here, we will briefly introduce the concept for the case of a square matrix. For a more detailed discussion, see [27]. The singular value decomposition of $m_1(\frac{1}{2})$, is given by:

$$m_1(\tfrac{1}{2}) = USV^\dagger, \tag{5.48}$$

where U and V are unitary matrices and S is a diagonal matrix, consisting of the singular values, which are real and non-zero. If the columns of U and V are denoted by u_i and v_i, and the singular values by s_i we get:

$$m_1(\tfrac{1}{2})v_i = s_i u_i. \tag{5.49}$$



(a) Lower surface state.

(b) Upper surface state

Figure 5.8: The density at $L = 30\,\text{Å}$ for the upper and lower surface states. Exactly as in the semi-infinite case we see that the lower state gets more localized in the beginning when increasing $k_{||}$, before spreading out again and going in to the bulk. The upper state just spreads in to the bulk directly, for a relatively small value of $k_{||}$.

If some $s_i = 0$, then the corresponding vector v_i is a solution to eq. (5.47), and if all $s_i \neq 0$, then there is only the trivial solution $\mathbf{C} = 0$. Therefore, we take out the vector corresponding to the smallest singular value, which will be an approximate solution to eq. (5.47). The smallest singular value would be zero, if we had used exact value for $E(k_{||})$. We assume that there is no other degeneracy than the one caused by TI invariance, and we only look for one solution to each block. This gives us the coefficients, and we can calculate the wave function. The other degenerate solution belonging to the lower block, is simply obtained using the TI operator on the found wave function.

The notion of surface state is not as clear cut as in the semi-infinite case, since all states have non-zero wave functions throughout the topological insulator. Here we could define a surface state, as a state with an energy in the bulk gap. According to that definition there is only one surface state for $L = 20\,\text{Å}$, while at $L = 40\,\text{Å}$ we see two surface states on figure 5.6. And as we see on figure 5.8, the surface states becomes bulk states when $k_{||}$ is increased, exactly as in the semi-infinite case.

The connection to the semi-infinite case can be made when L becomes large enough that the gap can be neglected. In this limit the surface state wave functions are almost zero in the center of the insulator as we see on figure 5.5 and we can make superpositions located on either surface of the material. For $k_{||} \neq 0$ we make superpositions of the degenerate TI partners:

$$\Psi_{k_x,k_y,n} \pm TI\Psi_{k_x,k_y,n} \tag{5.50}$$

which gives to new states located on each surface, as seen on figure 5.9. At the gamma point the TI partners are purely from the spin up block or the spin down block, and hence, they cannot cancel out. Therefore we need to combine states from both the upper and lower surface band with the same spin, to get a state localized on one surface. This is possible, since we are in the limit of large L and hence the gap is neglected, and there is 4 degenerate states at $k_{||} = 0$ (even/odd and spin up/down).

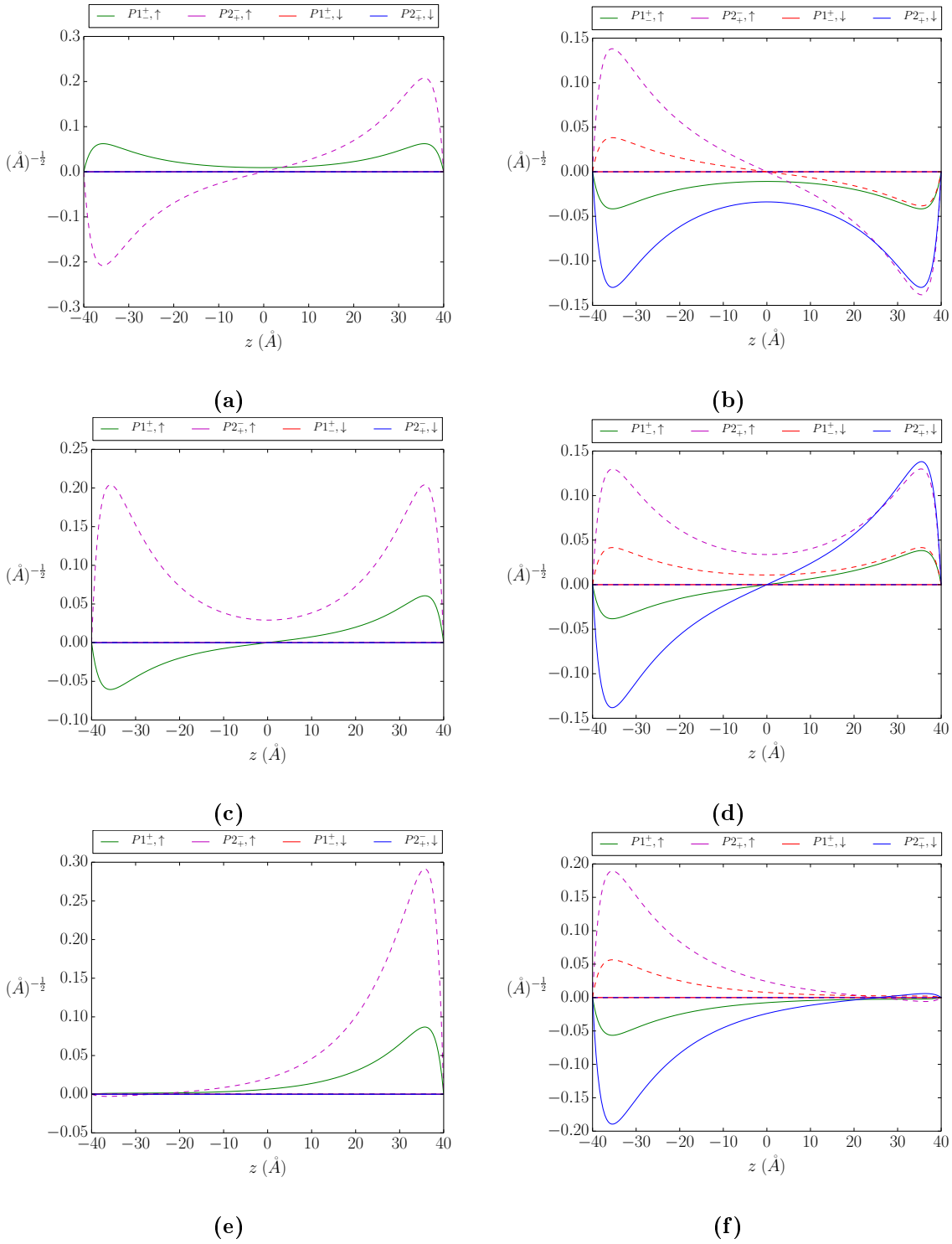


Figure 5.9: For $L = 80$ Å the gap is negligible, and we can make superpositions located on a single surface. (a) and (c) shows the even and odd spin up wave functions at the gamma point, and the superposition of the them, located on a single surface, is shown in (e). (b) and (d) shows the two degenerate eigenstates at $k_x = 0.01$ Å, for which all 4 components are nonzero. Their superposition, located on a single surface is shown in (e). If we had taken the difference instead of the sum, we would have gotten a state located at the opposite surface. The color denotes the 4 different components, in the basis of our model Hamiltonian. The solid lines are the real part, while dashed are the imaginary part.

5.3.1 Spin

We saw in the semi-infinite case that the surface states exhibits spin momentum locking. One could think that when we have an inversion symmetric system this is not the case, since:

$$TI\mathbf{S}TI = -\mathbf{S}, \tag{5.51}$$

and then the two doubly degenerate states have opposite spin, and hence can form superpositions giving spin states in any direction. But if we look at the spatial dependence of the spin we see a momentum dependence. And since we cannot measure each of the doubly degenerate states individually, we want to calculate the sum of the expectation values of the spin of the two TI partners. We only need to do this for $k_y = 0$. The rotational symmetry in the (x, y) plane, ensures that when you rotate into some other state with $k_y \neq 0$, the spin is just rotated along. Therefore, if we find spin-momentum locking at some $k_{||} = k_x$, then there must be spin momentum locking for all states with this magnitude of in plane momentum $k_{||}$.

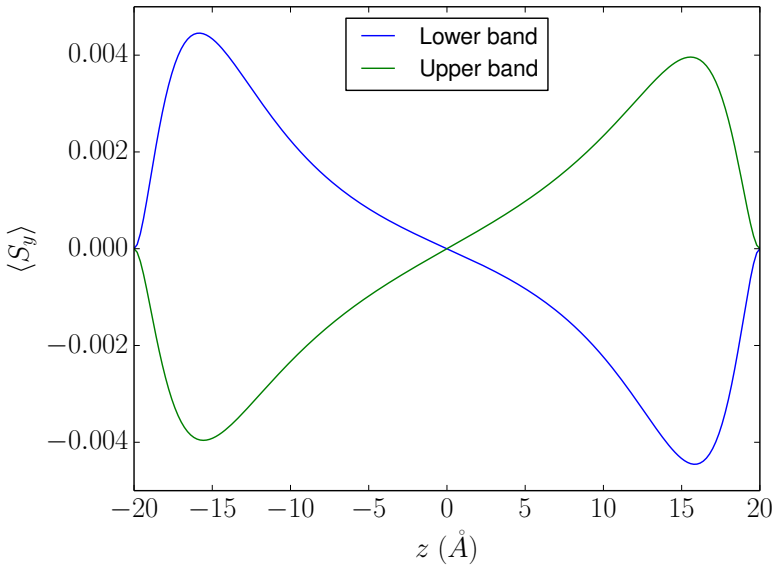


Figure 5.10: The spatial dependence of the expectation value of the spin in the y direction for a surface states in an insulator with $L = 40$ Å at $k_x = 0.001, k_y = 0$. The spin in the other directions is zero.

We see on figure 5.10 that the spin is opposite on the two surfaces. For the upper band, $\langle S_y \rangle \propto k_x$ at the top surface. Therefore, the top surface shows the same vorticity of the spin-momentum locking as the semi-infinite insulator, in section 4.4.2. The bottom surface has the opposite vorticity. This fact is due to the inversion symmetry; The inversion operator interchanges the two surfaces, and takes $k_x, k_y \to -k_x, -k_y$. Therefore, since the spin is unchanged, the vorticity of the spin-momentum locking must be reversed.

The dependence of the spin expectation value on L and $k_{||}$, follows completely from the general dependence of the wavefunctions. When L increases the two peaks stays close to the surface, while the value in the middle goes to zero, and when $k_{||}$ increases they extent further into to the material.

5.4 Local density of states

Again we will calculate the local density of states, here including both surface states and bulk states.

$$\rho(z,\omega) = \sum_{k_x,k_y,n} \frac{2}{A} |\Psi_{k_x,k_y,n}(z)|^2 \delta(\omega - E_n(k_||)) \tag{5.52}$$

n is an index indicating the band, which can be both surface and bulk bands. The factor of 2 is because of the double degeneracy due to TI symmetry. Since we consider a system, which is infinite in the x, y plane we take the limit $A \to \infty$ turning the sums over momenta into integrals:

$$\rho(z,\omega) = \frac{1}{(2\pi)^2} \sum_n \int_{\text{FBZ}} \mathrm{d}k_x \mathrm{d}k_y |\Psi_{k_x,k_y,n}(z)|^2 \delta(\omega - E_n(k_{||})) \tag{5.53}$$

The integral is in principle over the first brillouin zone, but since we only consider low energies we will have som cut-off. By the symmetry under rotations around the z axis, we can change to polar coordinates and integrate out the angle:

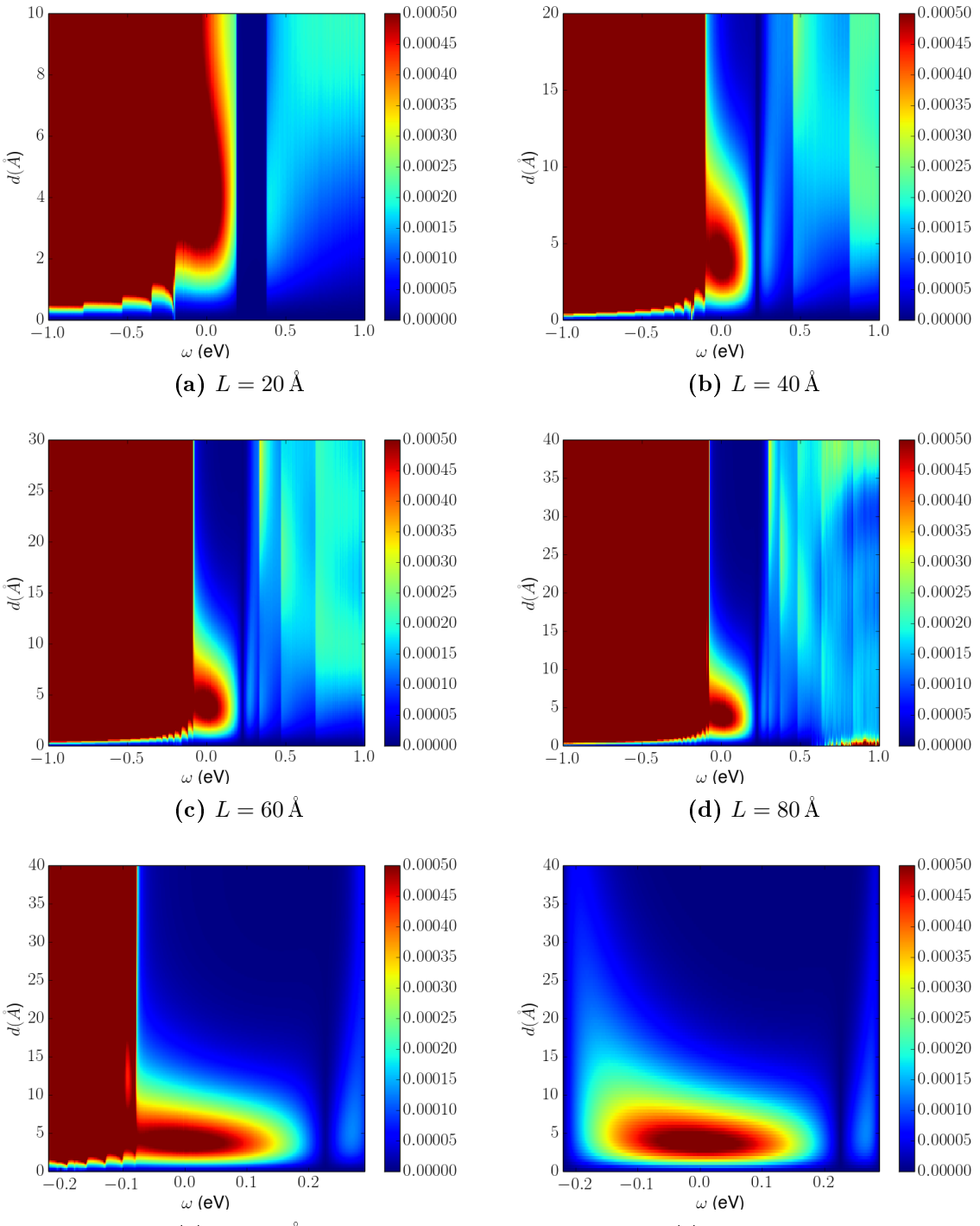
$$\rho(z,\omega) = \frac{1}{2\pi} \sum_n \int_0^{k_c} \mathrm{d}k_{||} k_{||} |\Psi_{k_x,k_y,n}(z)|^2 \delta(\omega - E_n(k_{||})) \tag{5.54}$$

For a given $k_{||}$, we can calculate the eigenenergies and the corresponding wave functions numerically. We will approximate the $k_{||}$ integral with a sum, and therefore we need to replace the delta function by some distribution with a finite width. We will use a Lorentz distribution $L(E)$, i.e.

$$\rho(z,\omega) = \frac{1}{2\pi} \sum_{k_{||}} \sum_n \delta k_{||} k_{||} |\Psi_{k_x,k_y,n}(z)|^2 L(\omega - E_n(k_{||})). \tag{5.55}$$

The $k_{||}$ is summed over equidistant values from 0 to some $k_{\max}$, with the spacing $\delta k_{||}$. The sum should in principle go to infinity, but since we are only interested in a finite energy range, we can terminate the sum at $k_{\max}$, where all $E_n(k_{\max})$ are out of this range. The width of the Lorentz distribution has to be small enough that the details of the function does not get washed out, but big enough that that you do not see the individual $k_{||}$ points in the sum.

The LDOS are plotted in figure 5.11. We clearly see the energy gap for $L = 20\,\text{Å}$, throughout the topological insulator. For increased L we see the surface gap closing, but in the center of the insulator we see the bulk gap. The bulk bands are seen as vertical lines, throughout the material. Note the similarity of the LDOS for $L = 80\,\text{Å}$ and the semi-infinite case. The only difference is the large contribution from the bulk states for $L = 80\,\text{Å}$, which was not included in the semi-infinite case.



(a) $L = 20\,\text{Å}$ **(b)** $L = 40\,\text{Å}$

(c) $L = 60\,\text{Å}$ **(d)** $L = 80\,\text{Å}$

(e) $L = 80\,\text{Å}$ **(f)** Semi-infinite

Figure 5.11: The local density of states for various thicknesses, as a function of distance from the surface d and energy ω. Note that the color scale is the same for all the plots, with an upper cut-off, to see the surface states clearly. For the plots with finite L, the range of d is $\frac{L}{2}$, the other half is the same due to inversion symmetry. We see the large gap for $L = 20\,\text{Å}$ and at $L = 80\,\text{Å}$ the gap is almost disappeared. For comparison we show the LDOS of the surface states in the semi-infinite case in (f) and for $L = 80\,\text{Å}$ in (e), with the same range of d and ω. At the energies where the bands are flat (e.g. at an extremum point) we get a large contribution to the density of states. This explains the vertical lines in the LDOS, if we compare to 5.6.

Summary and discussion

In this thesis, we have investigated the 3D topological insulator Bi_2Se_3. We started by giving a detailed introduction to applications of group theory in physics. This gave us the necessary tools to proceed, and give a qualitative discussion on the electronic structure of Bi_2Se_3. With this at hand we constructed a low energy model, using the four bands closest to the Fermi level. This model was the starting point for the rest of the thesis, where we investigated this model under different boundary conditions.

In chapter 4, we gave a detailed derivation of the surface states of a semi-infinite topological insulator. We found the criteria for the existence of surface states. Surprisingly the surface states could be destroyed by a large particle-hole asymmetry. By calculating the spin expectation values of the surface states, we found a helical spin structure. The vorticity of the spin structure was found to be opposite above and below the Dirac point. The spin structure was independent of the model parameters, except the sign of the vorticity was determined by the sign of the parameter combination $A_1A_2B_1$. We considered two different sets of parameters, and surprisingly, these lead to two different vorticities. The spatial structure of the surface states showed a strong dependence on the in-plane momentum. For larger $k_{||}$, the surface states extended further into the material, with the expectation value of the distance to the surface eventually diverging, exactly when the surface band touches the $k_z = 0$ bulk band.

For a real sample, there is always two surfaces, and in chapter 5, we discussed the finite size effect of Bi_2Se_3. An energy band gap was induced, due to the coupling between the two surfaces. We saw an exponential decay of this gap, as expected, because of the exponential decay of the wave functions. At a thickness of six quintuple layers, approximately 60 Å, we found the gap to be below 0.01 Å, which is consistent with the experiment in [26] measuring no gap at this thickness. Therefore, for samples of 6 quintuple layers or more, we can safely use the semi-infinite boundary conditions. For 2-4 quintuple layers, the measured gap was larger than our theoretical prediction. There can be many reasons for this discrepancy. One reason could be higher order terms in k_z. Even though we consider the gap at the gamma point in the two-dimensional brillouin zone, the surface states are localized in the z direction. Therefore, we have no reason to expect a low order expansion in k_z to give exact results. These terms could in principle be included perturbatively. However, this would require to determine the model parameters of higher order terms.

When increasing the thickness, they find in [26] that the Dirac point increases in energy, and saturates at 20 quintuple layers. This effect is not seen in our theoretical model, where the gap closes exactly at the Dirac point for the semi-infinite system. But since the spectrum is measured only relative to the Fermi level, they argue that this is an effect of thickness dependent electron doping, changing the Fermi level.

Using the parameters of [12], we saw an oscillation of the gap on top of the decay. Experimental data from [26], show no sign of this oscillation. The closing of the gap, in the $L \to \infty$ limit, is guaranteed by time reversal symmetry. In contrast the gap closings at finite L are not protected by any symmetry. Therefore, it is reasonable to expect that these gap closings can

be destroyed by impurities. These gap closings are merely a result of the specific form of the wave functions and boundary conditions.

The spin structure of the surface states showed a position dependence, with opposite spin on opposite surfaces. This can be explained by the inversion symmetry of the crystal, which interchanges the two surfaces.

6.1 Outlook

There are many possibilities for further studies of three-dimensional topological insulators. Staying within this model, it could be interesting to consider other options for the boundary conditions, which according to [19], can change both the spectrum and even existence of surface states. In this article they a class of boundary conditions, where some linear combination of the wave function and its first derivative is required to vanish at the surface. Another option is to let the parameter M continuously change sign across the surface.

In [11], they include a magnetic field in a 2D model for the surface states. In the 2D model for thin films, both a magnetic field and a structural inversion asymmetry could be included. Then there is three different mechanisms to open a gap at the gamma point; Finite size effect, structural inversion asymmetry and magnetic field. It could be interesting to see how these affect the edge states of a thin film.

Here we have only considered surfaces, parallel to the atomic layers. In [17] they show, that a different spin structure arises for surfaces perpendicular to the atomic layers, by considering a semi-infinite topological insulator with hard-wall boundary conditions. It could be interesting to consider a finite system, with two surfaces not parallel to the atomic layers. However, it might be an experimental challenge to fabricate such a sample, due to the structure of the crystal.

Another conceptually interesting option, would be to consider a semi-classical approach to the surface states. Then it might be possible to calculate semi-classical orbits, similar to skipping orbits in a quantum hall system [28].

2D model for thin film Bi_2Se_3

In this appendix, we will show that the 2D model for thin films of 3D topological insulators, found in section 5.1.2 and in [16] is unitary equivalent to the BHZ model for HgTe quantum wells given in [5]

The 2D model is given by:

$$H_2D = E_0 - Dk_{||}^2 + \begin{pmatrix} M - Bk_{||}^2 & \tilde{A}_2 k_- & 0 & 0 \\ \tilde{A}_2^* k_+ & -M + Bk_{||}^2 & 0 & 0 \\ 0 & 0 & -M + Bk_{||}^2 & -\tilde{A}_2^* k_- \\ 0 & 0 & -\tilde{A}_2 k_+ & M - Bk_{||}^2 \end{pmatrix} \tag{A.1}$$

where we have defined a new parameter $M = \frac{\Delta}{2}$, to use the same notation as in the BHZ model. In the appendix of [16] they show that $\tilde{A}_2$ is either purely real or purely imaginary. If we for the real one define $A = \tilde{A}_2$ and for the imaginary one $A = -i\tilde{A}_2$ we have the two models:

$$H_R = E_0 - Dk_{||}^2 + \begin{pmatrix} M - Bk_{||}^2 & Ak_- & 0 & 0 \\ Ak_+ & -M + Bk_{||}^2 & 0 & 0 \\ 0 & 0 & -M + Bk_{||}^2 & -Ak_- \\ 0 & 0 & -Ak_+ & M - Bk_{||}^2 \end{pmatrix} \tag{A.2}$$

$$H_I = E_0 - Dk_{||}^2 + \begin{pmatrix} M - Bk_{||}^2 & iAk_- & 0 & 0 \\ -iAk_+ & -M + Bk_{||}^2 & 0 & 0 \\ 0 & 0 & -M + Bk_{||}^2 & iAk_- \\ 0 & 0 & -iAk_+ & M - Bk_{||}^2 \end{pmatrix} \tag{A.3}$$

These two models are unitary equivalent since:

$$H_R = U_1 H_I U_1^\dagger \tag{A.4}$$

for

$$U_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & i & 0 & 0 \\ 0 & 0 & i & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{A.5}$$

The H_R can be transformed into the BHZ model by interchanging basis states 3 and 4, or

equivalent by:

$$H_{BHZ} = E_0 - Dk_{||}^2 + \begin{pmatrix} M - Bk_{||}^2 & Ak_- & 0 & 0 \\ Ak_+ & -M + Bk_{||}^2 & 0 & 0 \\ 0 & 0 & -M + Bk_{||}^2 & -Ak_- \\ 0 & 0 & -Ak_+ & M - Bk_{||}^2 \end{pmatrix} = U_2 H_R U_2^\dagger \quad \text{(A.6)}$$

where

$$U_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix} \quad \text{(A.7)}$$

So both case I and case II are unitary related to the BHZ model.

A.1 Edge states in the BHZ model

$$H_{BHZ} = \begin{pmatrix} h(\mathbf{k}) & 0 \\ 0 & h^*(-\mathbf{k}) \end{pmatrix} \quad \text{(A.8)}$$

Since the upper block is simply the time-reversal of the lower block we will focus on solving the upper.

$$h(k) = \varepsilon(k_{||}) + \sigma \cdot \mathbf{d} \quad \text{(A.9)}$$

where $\varepsilon(k_{||}) = C - Dk_{||}^2$ and $\mathbf{d} = (Ak_x, Ak_y, M - Bk_{||}^2)$. For convenience we just put $C = 0$ (it just defines the zero point of the energy). Now we seek solutions to a semi-infinite plane, in the region $y \leq 0$, with hardwall boundary conditions i.e.

$$\Psi(y = 0) = 0. \quad \text{(A.10)}$$

To ensure that the wavefunction is normalizable we also impose:

$$\Psi(y \to -\infty) = 0. \quad \text{(A.11)}$$

We use the ansatz $\Psi = \psi_\lambda e^{y\lambda}$. Then we get the eigenvalue equation:

$$\begin{pmatrix} M - D_+(k_x^2 - \lambda^2) - E & A(k_x - \lambda) \\ A(k_x + \lambda) & -M - D_-(k_x^2 - \lambda^2) - E \end{pmatrix} \psi_\lambda = 0, \quad \text{(A.12)}$$

where $D_\pm = D \pm B$. This gives the secular equation:

$$0 = (M - D_+(k_x^2 - \lambda^2) - E)(-M - D_-(k_x^2 - \lambda^2) - E) - A^2(k_x^2 - \lambda^2) \tag{A.13}$$
$$= E^2 - M^2 + D_+D_-(k_x^2 - \lambda^2)^2 + (2MB + 2DE - A^2)(k_x^2 - \lambda^2) \tag{A.14}$$
$$\Leftrightarrow k_x^2 - \lambda^2 = -F \pm \sqrt{F^2 - Q} \tag{A.15}$$
$$\Leftrightarrow \lambda_{1,2} = \sqrt{k_x^2 + F \pm \sqrt{F^2 - Q}} \tag{A.16}$$

Where we defined $F = \frac{2MB+2DE-A^2}{2D_+D_-}$ and $Q = \frac{E^2-M^2}{D_+D_-}$. The boundary value at $y \to -\infty$ excludes the negative (outer) squareroot. The boundary value at $y = 0$ gives:

$$\psi_{\lambda_1} + \psi_{\lambda_2} = 0 \tag{A.17}$$

The eigenvectors can be written in two ways (which are equivalent if the secular equation is satisfied);

$$\psi_\lambda = \begin{pmatrix} A(\lambda - k_x) \\ M - D_+(k_x^2 - \lambda^2) - E \end{pmatrix} = \begin{pmatrix} M + D_-(k_x^2 - \lambda^2) + E \\ A(k_x + \lambda) \end{pmatrix}. \tag{A.18}$$

The equation (A.17) means that that ψ_{λ_1} and ψ_{λ_2} has to be linearly dependent i.e., the determinant of $(\psi_{\lambda_1}\psi_{\lambda_2})$ is equal to zero. Using the first form of (A.18) gives the equation:

$$0 = A(\lambda_1 - k_x)(M - D_+(k_x^2 - \lambda_2^2) - E) - A(\lambda_2 - k_x)(M - D_+(k_x^2 - \lambda_1^2) - E) \tag{A.19}$$
$$0 = D_+k_x(\lambda_1^2 - \lambda_2^2) + (\lambda_1 - \lambda_2)(M - E - D_+(k_x^2 + \lambda_1\lambda_2)) \tag{A.20}$$
$$0 = D_+k_x(\lambda_1 + \lambda_2) + (M - E - D_+(k_x^2 + \lambda_1\lambda_2)) \tag{A.21}$$
$$\frac{M - E}{D_+} = k_x^2 + \lambda_1\lambda_2 - k_x(\lambda_1 + \lambda_2) \tag{A.22}$$

And from the second form of (A.18) we obtain the equation:

$$\frac{M + E}{-D_-} = k_x^2 + \lambda_1\lambda_2 + k_x(\lambda_1 + \lambda_2) \tag{A.23}$$

The sum of these equations give:

$$\frac{D_-(M - E) - D_+(M + E)}{2D_+D_-} = k_x^2 + \lambda_1\lambda_2 \tag{A.24}$$
$$\Leftrightarrow \lambda_1\lambda_2 = -\frac{BM + DE}{D_+D_-} - k_x^2 \tag{A.25}$$

and the difference:

$$\frac{D_-(M - E) + D_+(M + E)}{2D_+D_-} = -k_x(\lambda_1 + \lambda_2) \tag{A.26}$$
$$\Leftrightarrow \lambda_1 + \lambda_2 = -\frac{DM + BE}{k_xD_+D_-} \tag{A.27}$$

From equation (A.25) we insert the form of the λ's and square to obtain:

$$(k_x^2 + F)^2 - F^2 + Q = \left(\frac{BM + DE}{D_+ D_-}\right)^2 + k_x^4 + 2k_x^2 \frac{BM + DE}{D_+ D_-} \tag{A.28}$$

$$\Leftrightarrow \frac{-A^2 k_x^2}{D_+ D_-} + \frac{E^2 - M^2}{D_+ D_-} = \left(\frac{BM + DE}{D_+ D_-}\right)^2 \tag{A.29}$$

$$\Leftrightarrow 0 = B^2 E^2 + 2MBDE + M^2 D^2 + D_+ D_- A^2 k_x^2 \tag{A.30}$$

$$\Leftrightarrow E = -\frac{MD}{B} \pm \frac{|A|}{|B|}\sqrt{-D_+ D_-}|k_x| \tag{A.31}$$

And we get a linear spectrum in k_x. If an edge state exists, then $\lambda_1 \lambda_2 > 0$, since λ_1 and λ_2 are either real and positive or complex conjugate partners. For $k_x = 0$ the energy is $E = -\frac{MD}{B}$, and inserting this in A.25 gives the condition:

Then we use equation (A.22) to rewrite the eigenvector as:

$$\psi_{\lambda_1} \propto \begin{pmatrix} 1 \\ \frac{M-E}{A(\lambda_1 - k_x)} + \frac{D_+}{A}(\lambda_1 + k_x) \end{pmatrix} = \begin{pmatrix} 1 \\ \frac{D_+}{A}(\lambda_2 - k_x) + \frac{D_+}{A}(\lambda_1 + k_x) \end{pmatrix} \tag{A.32}$$

$$= \begin{pmatrix} 1 \\ \frac{|A|}{A}\sqrt{-\frac{D_+}{D_-}} \end{pmatrix} \tag{A.33}$$

We see that it is independent of λ_1 (which it should be since it should be proportional to ψ_{λ_2}) and surprisingly also independent of k_x

$$0 < \lambda_1 \lambda_2 = -\frac{BM + DE}{D_+ D_-} = \frac{M}{B} \tag{A.34}$$

To find the edge state for the lower block we use that the hamiltonian is time reversal invariant:

$$\theta H(k)\theta^{-1} = -i\sigma_y K \begin{pmatrix} h(k) & 0 \\ 0 & h^*(-k) \end{pmatrix} i\sigma_y K = K \begin{pmatrix} h^*(-k) & 0 \\ 0 & h(k) \end{pmatrix} K = H(-k) \tag{A.35}$$

So we have that $\psi_\downarrow = \theta\psi_\uparrow$ is an eigenstate:

$$H(k)\psi_\uparrow = E_k^\uparrow \psi_\uparrow \tag{A.36}$$

$$\Leftrightarrow \theta H(k)\theta^{-1}\theta\psi_\uparrow = E_k^\uparrow \theta\psi_\uparrow \tag{A.37}$$

$$\Leftrightarrow H(-k)\psi_\downarrow = E_k^\uparrow \psi_\downarrow \tag{A.38}$$

So we have:

$$E_k^\downarrow = E_{-k}^\uparrow \tag{A.39}$$

And the total eigenstates are

$$\psi_\uparrow = e^{ik_x x}(e^{\lambda_1 y} - e^{\lambda_2 y}) \begin{pmatrix} 1 \\ \frac{|A|}{A}\sqrt{-\frac{D_+}{D_-}} \\ 0 \\ 0 \end{pmatrix}, \psi_\downarrow = e^{ik_x x}(e^{\lambda_1 y} - e^{\lambda_2 y}) \begin{pmatrix} 0 \\ 0 \\ 1 \\ \frac{|A|}{A}\sqrt{-\frac{D_+}{D_-}} \end{pmatrix} \tag{A.40}$$

with eigenenergies:

$$E_k^\sigma = -\frac{MD}{B} \pm \frac{|A|}{B}\sqrt{-D_- D_+}k_x \tag{A.41}$$

with plus sign for spin up and minus sign for spin down.

The eigenstates for H_R are then $U_2\psi_{BHZ}$

$$\psi_1 = e^{ik_x x}(e^{\lambda_1 y} - e^{\lambda_2 y}) \begin{pmatrix} 1 \\ \frac{|A|}{A}\sqrt{-\frac{D_+}{D_-}} \\ 0 \\ 0 \end{pmatrix}, \psi_2 = e^{ik_x x}(e^{\lambda_1 y} - e^{\lambda_2 y}) \begin{pmatrix} 0 \\ 0 \\ \frac{|A|}{A}\sqrt{-\frac{D_+}{D_-}} \\ 1 \end{pmatrix} \tag{A.42}$$

But in this basis the two first basis states have different spins, so both eigenstates are superpositions of spin up and spin down.

And for case II (H_I) we get ($U_1^\dagger \psi_R$)

$$\psi_1 = e^{ik_x x}(e^{\lambda_1 y} - e^{\lambda_2 y}) \begin{pmatrix} -i \\ \frac{|A|}{A}\sqrt{-\frac{D_+}{D_-}} \\ 0 \\ 0 \end{pmatrix}, \psi_2 = e^{ik_x x}(e^{\lambda_1 y} - e^{\lambda_2 y}) \begin{pmatrix} 0 \\ 0 \\ i\frac{|A|}{A}\sqrt{-\frac{D_+}{D_-}} \\ 1 \end{pmatrix} \tag{A.43}$$

Bibliography

[1] K. von Klitzing, G. Dorda, and M. Pepper. New method for high-accuracy determination of the fine-structure constant based on quantized hall resistance. *Phys. Rev. Lett.*, 45(494), 1980.

[2] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov. Electric field effect in atomically thin carbon films. *Science*, 306(5696), 2004.

[3] C. L. Kane and E. J. Mele. Quantum spin hall effect in graphene. *Phys. Rev. Lett.*, 95(226801), 2005.

[4] Y. Ando. Topological insulator materials. *J. Phys. Soc. Jpn.*, 82(102001), 2013.

[5] B. A. Bernevig, T. L. Hughes, and S. Zhang. Quantum spin hall effect and topological phase transition in hgte quantum wells. *Science*, 314(1757), 2006.

[6] M. König, S. Wiedmann, C. Brüne, A. Roth, H. Buhmann, L. W. Molenkamp, X. Qi, and S. Zhang. Quantum spin hall insulator state in hgte quantum wells. *Science*, 318(5851), 2007.

[7] Y. Xia, D. Qian, L. Wray, A. Pal, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan. Observation of a large-gap topological-insulator class with a single dirac con on the surface. 5(398), 2009.

[8] Michael Tinkham. *Group Theory and Quantum Mechanics*. McGraw-Hill Book Company, 1964.

[9] S. A. Werner, R. Colella, A. W. Overhauser, and C. F. Eagen. Observation of the phase shift of a neutron due to precession in a magnetic field. *Phys. Rev. Lett.*, 35(1053), 1975.

[10] Mildred S. Dresselhaus, Gene Dresselhaus, and Ado Jorio. *Group Theory: Application to the physics of condensed matter*. Springer, 2008.

[11] C. Liu, X. Qi, H. Zhang, X. Dai, Z. Fang, and S. Zhang. Model hamiltonian for topological insulators. *Phys. Rev. B*, 82(045122), 2010.

[12] H. Zhang, C. Liu, X. Qi, X. Dai, Z. Fang, and S. Zhang. Topological insulators in Bi_2Se_3, Bi_2Te_3 and Sb_2Te_3 with a single dirac cone on the surface. *Nat. Phys.*, 5(438), 2009.

[13] W. Zhang, R. Yu, H. Zhang, X. Dai, and Z. Fang. First-principles studies of the three-dimensional strong topological insulators Bi_2Te_3, Bi_2Se_3 and Sb_2Te_3. *New J. Phys*, 12(065013), 2010.

[14] S. K. Mishra, S. Satpathy, and O. Jepsen. Electronic structure and thermoelectric properties of bismuth telluride and bismuth selenide. *J. Phys.: Condens. Matter*, 9(461), 2010.

[15] J. Sakurai and J. Napolitano. *Modern Quantum Mechanics*. Pearson, 1994.

[16] W. Shan, H. Lu, and S. Shen. Effective continuous models for surface states and thin films of three-dimensional topological insulators. *New J. Phys.*, 12(043048), 2010.

[17] P. G. Silvestrov, P. W. Brouwer, and E. G. Mishchenko. Spin and charge structure of the surface states in topological insulators. *Phys. Rev. B*, 86(075302), 2012.

[18] Roland Winkler. *Spin-orbit Coupling Effects in Two-Dimensional Electron and Hole Systems*. Springer, 2003.

[19] V. V. Enaldiev, I. V. Zagorodnev, and V. A. Volkov. Boundary conditions for effective hamiltonian and surface states in 2d and 3d topological insulators. arXiv:1407.0945, 2014.

[20] Angle-resolved photoemission spectroscopy. Angle-resolved photoemission spectroscopy — *Wikipedia, The Free Encyclopedia*, 2014. [Online; accessed 22-July-2014].

[21] T. Schlenk, M. Bianchi, M. Koleini, A. Eich, O. Pietzsch, T. O. Wehling, T. Frauenheim, A. Balatsky, J.-L. Mi, B. B. Iversen, J. Wiebe, A. A. Khajetoorians, P. Hofmann, and R. Wiesendanger. Controllable magnetic doping of the surface state of a topological insulator. *Phys. Rev. Lett.*, 110(126804), 2013.

[22] Henrik Bruus and Karsten Flensberg. *Many-Body Quantum Theory in Condensed Matter Physics*. Oxford University Press, 2004.

[23] H. Lu, W. Shan, W. Yao, Q. Niu, and S. Shen. Massive dirac fermions and spin physics in an ultrathin film of topological insulator. *Phys. Rev. B*, 81(115407), 2010.

[24] C. Liu, H. Zhang, B. Yan, X. Qi, T. Frauenheim, X. Dai, Z. Fang, and S. Zhang. Oscillatory crossover from two-dimensional to three-dimensional topological insulators. *Phys. Rev. B*, 81(041307), 2010.

[25] J. Linder, T. Yokoyama, and A. Sudbø. Anomalous finite size effects on surface states in the topological insulators Bi_2Se_3. *Phys. Rev. B*, 80(205401), 2009.

[26] Y. Zhang, K. He, C. Chang, C. Song, L. Wang, X. Chen, J. Jia, Z. Fang, X. Dai, W. Shan, S. Shen, Q. Niu, X. Qi, S. Zhang, X. Ma, and Q. Xue. Crossover of the three-dimensional topological insulator Bi_2Se_3 to the two-dimensional limit. *Nat. Phys.*, 6(584), 2010.

[27] K. F. Riley, M. P. Hobson, and S. J. Bence. *Mathematical Methods for Physics and Engineering*. Cambridge University Press, 2006.

[28] Gilles Montambaux. Semiclassical quantization of skipping orbits. *Eur. Phys. J. B*, 78(215), 2011.