UTRECHT UNIVERSITY
INSTITUTE FOR THEORETICAL PHYSICS
MSc THEORETICAL PHYSICS

Topological Wannier Excitons in Bismuth Chalcogenide Nanosheets

Author
Lucas MAISEL LICERÁN

Supervisors
Henk STOOF
Lars FRITZ

August, 2022

Abstract

Current optoelectronic technologies depend crucially on insulators and semiconductors, whose response to photoexcitation at room temperature is largely dominated by bound electron-hole pairs known as excitons. While the physics of excitons in regular materials are very well understood, knowledge of their behavior in topological insulators is still severely limited. These novel phases of matter are receiving much attention due to their excellent prospects for energy-efficient electronics, (pseudo)spintronics devices and quantum-information processing. Therefore, a full understanding of light-matter interactions in these materials requires the investigation of physical and topological properties of bulk excitons.

In this work, we analyze the topology and dispersion of bulk Wannier excitons in thin nanosheets of bismuth selenide (Bi_2Se_3), a prototypical three-dimensional topological insulator. We find that excitons inherit the topology of the underlying electronic band structure, quantified by the winding numbers of the constituent electron and hole pseudospins as a function of the total exciton momentum. We also show that every s-wave exciton state consists of a nonchiral doublet of degenerate states with quadratic dispersion for low momenta, as well as a chiral doublet with one linear mode and one quadratic mode. We derive an effective model for the chiral excitons and consider their topological properties. Furthermore, we analyze the many-body screening due to the coherent surface states on the effective bulk electron-hole interaction, and consider the effect of surface plasmons. Our study is backed up by self-consistent numerical calculations and paints a complete picture of bulk excitons in quasi-two-dimensional topological materials with a band inversion at the Γ point. This picture can now in principle be used for the investigation of interactions between excitons themselves and other excitations such as plasmons and photons.

Contents

1

Introduction

In this work we will study excitons in a topological insulator. This will require familiarity with the physics and mathematical treatment of these concepts. The purpose of this preliminary chapter is precisely to provide the reader with the appropriate background. Several of the concepts presented in this chapter will not be needed to a very deep extent. However, we have chosen to provide some level of detail to put them into the wider context of the most relevant literature. We begin by introducing some basic notions of the mathematical branch of topology before moving on to the concept of a topological insulator. We then give some physical examples of such systems and introduce the particular material we shall be concerned with in the remainder of this thesis. Finally, we briefly discuss the concept of excitons and introduce some basic terminology ubiquitous in the literature.

1.1 Topology in condensed matter

The mathematical branch of topology consists in the study of continuous deformations between functions. Many of the physical phenomena that will be explored in this work can ultimately be traced back to nontrivial topological properties of the system. Therefore, in this section we present a brief summary of the main ideas behind this powerful tool. We will favor intuition over rigor and closely follow the physically motivated exposition of Refs. [1, 2, Ch. 1]. A thorough review of topological concepts in condensed-matter physics can be found in Ref. [3].

1.1.1 Homotopy

Let $f, g \colon M \to X$ be continuous functions between two topological spaces M and X. In the context of condensed-matter physics, X is often called the *order parameter space*. A *homotopy* between f and g is a continuous map $H \colon M \times [0,1] \to X$ such that $H(x,0) = f(x)$ and $H(x,1) = g(x)$ for all $x \in M$. As a simple example let $M = S^1$ in this section and assume that the mappings start and end at the same base point $x_0 \in X$, so that they constitute loops. Then f and g are said to be *homotopically equivalent* if there exists a homotopy between them, that is, if their respective loops can be continuously deformed into each other. We denote this equivalence by $f \sim g$. A map that is homotopic to the constant map x_0 is called *null-homotopic*

or *topologically trivial*. The corresponding loop can be continuously deformed to a point. Fig. 1.1 shows several examples of loops. Since homotopy is an equivalence relation, the set of loops in X can be divided into disjoint *homotopy classes*. The equivalence class of f is denoted by $[f]_{x_0}$ (we will drop the subscript x_0 when no confusion arises), and when $f \sim g$ their equivalence classes are identical, $[f] = [g]$. The homotopy class of a null-homotopic map f is denoted by $[f] = [e]$, where e denotes the constant map x_0.

(a)   **(b)**   **(c)**   **(d)**   **(e)**

$$[f] = [e] \qquad [f] \neq [e] \qquad [f] = [g] \qquad [f] \neq [g] \qquad P_c([f]) = [c \circ f \circ c^{-1}]$$
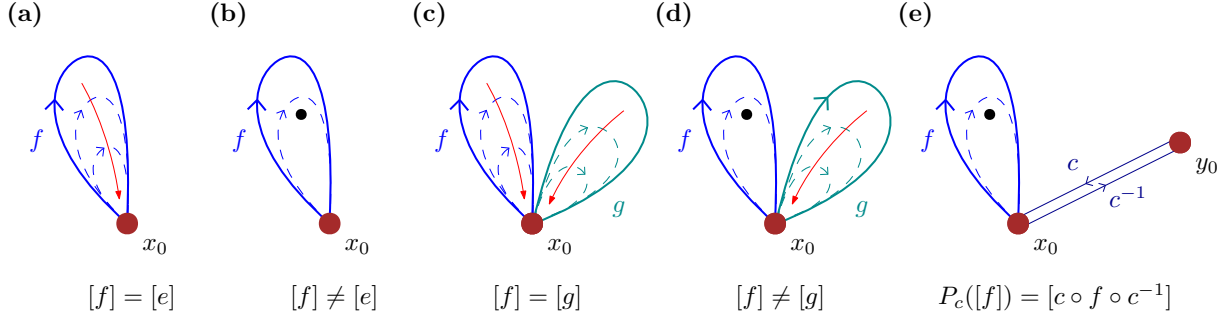
Figure 1.1. Examples of homotopies. **(a)** A null-homotopic loop, which can be continuously deformed to a point. The homotopy class of f then equals that of the constant map x_0, $[f] = [e]$. **(b)** If the target space has a "hole" (symbolized by a black point), there exist maps which cannot be deformed to x_0 and thus $[f] \neq [e]$. **(c)** Two homotopically equivalent maps, which can be continuously deformed into each other and thus $[f] = [g]$. **(d)** If the target space has holes then there exist homotopically inequivalent loops. **(e)** When the target space is path-connected, there exists a bijection between the sets of homotopy classes of any two points $x_0, y_0 \in X$. It is defined via the map $P_c([f]_{x_0}) = [c \circ f \circ c^{-1}]_{y_0}$, where c is any path between y_0 and x_0. Hence, there exists an isomorphism connecting the fundamental groups with different base points, $\pi_1(X, x_0) \cong \pi_1(X, y_0)$. Figure adapted from Ref. [1].

Two maps with the same base point x_0 can be concatenated to form a new loop $f \circ g$. This operation, in turn, defines a concatenation operation on homotopy classes via $[f] * [g] \equiv [f \circ g]$. It is straightforward to show that with this operation the set of homotopy classes becomes a group, denoted the *first homotopy group* or the *fundamental group* of X at x_0 and denoted by $\pi_1(X, x_0)$. Suppose now that the target space is path-connected, i.e., there exists a continuous path between any two points $x_0, y_0 \in X$. Then there exists a trivial isomorphism between fundamental groups at different base points, $\pi_1(X, x_0) \cong \pi_1(X, y_0)$ for all $x_0, y_0 \in X$. This isomorphism sends $[f]_{x_0} \to [c \circ f \circ c^{-1}]_{y_0}$, where $c: X \to X$ is any path from y_0 to x_0 [cf. Fig. 1.1 (e)]. Hence, in the path-connected case we can simply talk about the fundamental group $\pi_1(X)$.

What has been explained above for the case when the base manifold is S^1 can be generalized to maps between arbitrary topological spaces. It is not required to talk about loops in order to define the homotopy classes of a given target space. Furthermore, the set of homotopy classes of a map $S^n \to X$ always admits a group structure. The corresponding group $\pi_n(X)$ is called the *nth homotopy group* of X. All groups $\pi_n(X)$ with $n \geq 2$ are Abelian.

1.1.2 Winding numbers

We have now seen that mutually deformable mappings form equivalence classes. It is possible to define a quantity that serves to identify individual classes. This quantity is called the *degree* of the map f, $\deg f$, or also the *winding number w*. Of course, this means that configurations with different winding numbers cannot be continuously deformed into each other. It turns out that the winding numbers provide a bijection of the homotopy classes into the integers, as originally shown by Hopf. Two functions f and g are homotopically equivalent if and only if $\deg f = \deg g$, and the mapping is topologically trivial when $\deg f = 0$.

The most general expression for the winding number of f is given by

$$\deg f = \frac{\int_M f^*(\Omega)}{\int_X \Omega},\tag{1.1}$$

where Ω is a differential n-form defined on the target manifold X of dimension n, and $f^*(\Omega)$ denotes its pullback by f onto the base manifold M of dimension m [3, Ch. 5, 4, Ch. 2]. For concreteness consider the case $X \cong S^n$ and define local coordinates $\{x_i\}_{i=1}^m$ on M and $\{X_i\}_{i=1}^n$ on X. In this case Ω is most conveniently chosen as the surface volume form with respect to the metric induced from $\mathbb{R}^{n+1}$. Explicitly,

$$\Omega = \frac{1}{n!}\varepsilon_{\mu_0\mu_1\ldots\mu_n} X^{\mu_0}\, \mathrm{d}X^{\mu_1} \wedge \cdots \wedge \mathrm{d}X^{\mu_n},\tag{1.2}$$

because then $\mathrm{d}\Omega = (n+1)\omega$ with ω the standard volume form on $\mathbb{R}^{n+1}$. It follows from Stokes' theorem that $\int_{S^n}\Omega = (n+1)\int_{B^{n+1}}\omega = (n+1)V_{n+1}$, where B^{n+1} is an $(n+1)$-dimensional unit ball with boundary $\partial B^{n+1} = S^n$ and volume V_{n+1}. The pullback of Ω by f is

$$f^*(\Omega) = \frac{1}{n!}\varepsilon_{\mu_0\mu_1\ldots\mu_n} f^{\mu_0} \frac{\partial f^{\mu_1}}{\partial x^{\nu_1}} \cdots \frac{\partial f^{\mu_n}}{\partial x^{\nu_n}}\, \mathrm{d}x^{\nu_1} \wedge \cdots \wedge \mathrm{d}x^{\nu_n}.\tag{1.3}$$

Note that f^μ is shorthand for $(\psi \circ f \circ \varphi^{-1}(x))^\mu$, where $\varphi\colon M \to \mathbb{R}^m$ and $\psi\colon X \to \mathbb{R}^n$ are the local coordinate maps on M and X, respectively.

Physically speaking, one of the most interesting scenarios arises when the base manifold and the target manifold are m- and n-dimensional spheres, respectively. In the context of magnetism, for instance, one often works with a magnetization field $\boldsymbol{m}(\boldsymbol{x})$ in real space which can be mapped onto a sphere via stereographic projection[1]. Each point on the sphere is therefore assigned certain value of $\boldsymbol{m}$. Furthermore, there are different possibilities for the allowed spin orientations, i.e., for the spin-space dimensionality. Depending on the system, the spins may be allowed to only take two values (up or down, as in the case of Ising spins), be confined to a plane, or point in any direction in three-dimensional space. These correspond to $X \cong S^0$, S^1, and S^2, respectively. Thus, if m is the dimensionality of the space occupied by the spins and n is the dimensionality of the sphere traced by all their possible orientations, we are led to consider maps $f\colon S^m \to S^n$.

When the base and target manifolds have the same dimension $n \geq 1$, the homotopy group is simply

$$\pi_n(S^n) \cong \mathbb{Z},\tag{1.4}$$

and thus Abelian. It follows that each homotopy class can be characterized by an integer. This integer is nothing but the degree or winding number of the map, and it counts how many times the base manifold is wrapped around the target manifold under the map. We note that the case $n = 0$ is a slight exception to the above, and it is found that $\pi_0(S^0) \cong \mathbb{Z}_2$. Also, in the case $0 < m < n$ it is easy to see that $\pi_m(S^n) \cong \{e\}$ (the trivial group), because any mapping can be deformed to a point in the presence of an additional dimension in the target space. For instance, a map $f\colon S^1 \to S^2$ maps a circle onto a sphere, and one can always pull the circle back to a point by dragging each of its points towards the north pole. Finally, the case $m > n$ is nontrivial and will not be considered here.

To give a concrete physical example we consider a magnetization texture, or *skyrmion*, arising from spherical spins. The target space is S^2 and a point on the 2-sphere can be parametrized as $\boldsymbol{m} = (m^1, m^2, m^3)$ under the constraint $|\boldsymbol{m}| = 1$. The base manifold M is two-dimensional and can be parametrized by coordinates (x^1, x^2). These may refer to arbitrary curvilinear

---

[1]Note that this requires that the magnetization field assumes a constant value at infinity, because all points with $|\boldsymbol{x}| \to \infty$ get mapped to the north pole of the sphere.

coordinates such as, e.g., spherical ($M \cong S^2$), polar or Cartesian ($M \cong \mathbb{R}^2$) coordinates. The S^2 winding number then takes the form

$$w_{S^2} = \frac{1}{8\pi} \int_M \varepsilon_{abc} \, m^a \frac{\partial m^b}{\partial x^i} \frac{\partial m^c}{\partial x^j} \, dx^i \wedge dx^j = \frac{1}{4\pi} \int_M \boldsymbol{m} \cdot \left(\frac{\partial \boldsymbol{m}}{\partial x^1} \times \frac{\partial \boldsymbol{m}}{\partial x^2} \right) dx^1 \, dx^2. \tag{1.5}$$

If we use the usual representation $\boldsymbol{m} = (\sin\theta \sin\phi, \sin\theta \cos\phi, \cos\theta)$ in spherical coordinates, then this becomes

$$w_{S^2} = \frac{1}{4\pi} \int_M \sin\theta \left(\frac{\partial \theta}{\partial x^1} \frac{\partial \phi}{\partial x^2} - \frac{\partial \theta}{\partial x^2} \frac{\partial \phi}{\partial x^1} \right) dx^1 \, dx^2. \tag{1.6}$$

1.1.3 Topological robustness

We have now seen that skyrmions can be classified according to their winding number due to the relation $\pi_n(S^n) \cong \mathbb{Z}$. Textures with different winding numbers are not homotopically equivalent and thus cannot be continuously deformed into each other. This introduces a natural notion of *topological stability* or *robustness* in a magnetic system. Fig. 1.2 shows different trivial and nontrivial skyrmion configurations. In an idealized case, a skyrmion with nonzero winding number is impossible to disentangle to a topologically trivial state due to the need to surpass an infinite energy barrier. In practice this barrier is not infinite due to the discreteness of the lattice. Furthermore, the finiteness of experimental samples sometimes makes it possible to inject or eject topological defects through the edges. For this reason, real skyrmions are not entirely indestructible.
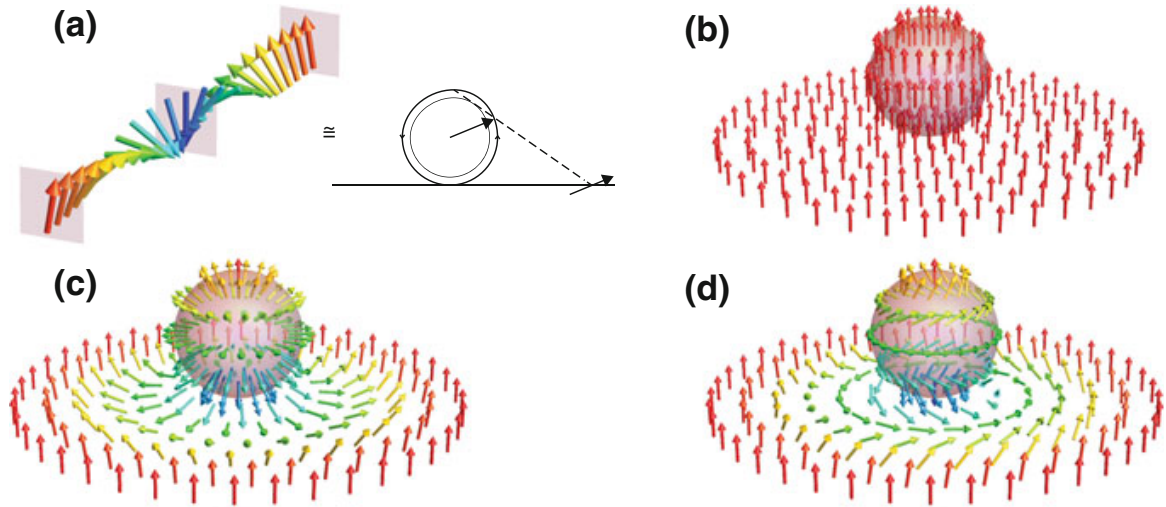
Figure 1.2. Examples of magnetization textures or skyrmions in a two-dimensional magnetic spin system. **(a)** A magnetization texture with constant value at infinity can be mapped to a sphere via stereographic projection, providing a means to observe the nontrivial winding of the skyrmion. **(b)** A topologically trivial skyrmion configuration with zero winding number, for which all the spins on the sphere point in the same direction. **(c)** A skyrmion with a winding number $|w_{S^2}| = 1$ is such that the spins wind once around the sphere when crossing the xy-plane in any direction. **(d)** Another skyrmion with unit winding number, which is homotopically equivalent to that in (c). The skyrmions in (c) and (d) are topologically distinct from the trivial magnetization texture in (b). Source: Ref. [2, Ch. 1].

Even though this notion of topological robustness is most easily understood for a magnetic system, it can be readily generalized to many other areas of physics. In the problem of describing the band structure of a semiconductor at low energies, is often the case that a certain vector-valued parameter formally plays the role of a magnetic field. For instance, one often encounters

the Hamiltonian

$$H(\boldsymbol{k}) = \boldsymbol{d}(\boldsymbol{k}) \cdot \boldsymbol{\sigma}, \tag{1.7}$$

where the components of $\boldsymbol{\sigma}$ are the three Pauli matrices spanning the Lie algebra $\mathfrak{su}(2)$. Often $\frac{1}{2}\boldsymbol{\sigma}$ describes the spin of a fermion, although the above Hamiltonian may also describe other (more abstract) degrees of freedom. The vector $\boldsymbol{d}(\boldsymbol{k})$, with $\boldsymbol{k}$ the crystal momentum, plays the role of a magnetic field coupling to these degrees of freedom. The topological properties of the resulting eigenstates are then fully determined by $\boldsymbol{d}(\boldsymbol{k})$ and its potentially nontrivial winding around the unit sphere. One subtlety here is that in an n-dimensional system the crystal momentum does not lie on S^n or $\mathbb{R}^n$. Rather, the space of all allowed momenta is isomorphic to the torus $\mathbb{T}^n$ owing to the periodicity of the Brillouin zone. In this context, then, one is led to consider the problem of finding the set of homotopy classes of maps $\mathbb{T}^2 \to X$, with X a certain *classifying space*. As we will see, identifying which classifying spaces are physically relevant and finding the corresponding homotopy classes ultimately leads to an elegant classification of fermionic condensed-matter systems.

1.2   Topological insulators

The quantum-mechanical band theory of solids developed in the 20th century provides a way to describe the electronic structure of metallic and insulating states [5, 6]. The translational symmetry of the crystal allows for the definition of a periodic Brillouin zone, a collection of crystal momenta $\boldsymbol{k}$ according to which all physical states can be classified. The band structure is described by a Bloch Hamiltonian $H(\boldsymbol{k})$, whose eigenstates are the Bloch states $|u_n(\boldsymbol{k})\rangle$, defined in a single unit cell. The corresponding eigenvalues $E_n(\boldsymbol{k})$ define energy bands which collectively make up the band structure of the crystal. An insulator is a system where the filled valence bands and the empty conduction bands are separated by a nonzero energy gap everywhere within the Brillouin zone. The simplest such state is the electrically inert *atomic insulator*, in which electrons are bound to atoms in closed shells. Interaction between the atoms leads to covalent bonding and a smaller energy gap, resulting in a semiconductor. Even though an atomic insulator and a semiconductor differ in the details of their band structure, there is a sense in which both can be regarded as the same phase. One can imagine tuning the Hamiltonian so as to continuously interpolate between the two without closing the energy gap at any point of the process. In this way one obtains a topological equivalence between different insulating states. In this sense, then, all conventional insulators and semiconductors are equivalent to each other and in fact equivalent to the vaccum of Dirac's relativistic theory of the electron.

The question that arises now is whether such a continuous deformation is always possible between two arbitrary insulators. In other words, are all insulating states topologically equivalent to the Dirac vacuum? The answer is no, and the counterexamples correspond to fascinating states of matter known as *topological insulators*. A topological insulator is simply a state of matter for which the deformation to the atomic insulator cannot be performed without closing the energy gap somewhere in the Brillouin zone at some point of the process. In this section, which closely follows the exposition of Refs. [7, 8], we give an introduction to the physics and classification of these novel phases of matter.

1.2.1   Antiunitary symmetries

In order to better understand how to classify topological insulators one needs to be familiar with some special symmetries that occur in condensed-matter systems. Here we give an overview of how these symmetries are realized in nonsuperconducting fermionic systems, which is the relevant setting for our purposes.

Let $\{\hat{\psi}_I, \hat{\psi}_I^\dagger\}_{I=1,\dots,N}$ be a set of fermionic creation and annihilation operators satisfying the anticommutation relations

$$\{\hat{\psi}_I, \hat{\psi}_J^\dagger\} = \delta_{IJ}. \tag{1.8}$$

The indices I and J stand for all possible combinations of the degrees of freedom of the particles, such as lattice site or spin. A general noninteracting fermion system is described by a second-quantized Hamiltonian $\hat{H}$. For a nonsuperconducting system, $\hat{H}$ is generically given by

$$\hat{H} = H_{IJ}\hat{\psi}_I^\dagger\hat{\psi}_J = \hat{\psi}^\dagger H \hat{\psi}. \tag{1.9}$$

Note that in the last expression we use matrix notation, with $\hat{\psi}$ being an N-component vector and H an $N \times N$ matrix corresponding to the single-particle (i.e., first-quantized) Hamiltonian.

Consider a transformation of the creation and annihilation operators, $\hat{\psi} \to \hat{\mathcal{U}}\hat{\psi}\hat{\mathcal{U}}^{-1}$, with $\hat{\mathcal{U}}$ some operator on the Fock-space. The transformation via $\hat{\mathcal{U}}$ corresponds to a symmetry of the system if it preserves both the anticommutation relations and the Hamiltonian,

$$\hat{\mathcal{U}}\{\hat{\psi}_I, \hat{\psi}_J^\dagger\}\hat{\mathcal{U}}^{-1} = \{\hat{\psi}_I, \hat{\psi}_J^\dagger\}, \tag{1.10a}$$

$$\hat{\mathcal{U}}\hat{H}\hat{\mathcal{U}}^{-1} = \hat{H}. \tag{1.10b}$$

According to Wigner's symmetry representation theorem, any symmetry transformation in quantum mechanics can be represented on a Hilbert space by an operator that is either linear and unitary, or antilinear and antiunitary. For our purposes we will be mainly interested in analyzing all possible *antiunitarily realized* symmetries, i.e., those for which either the first- or second-quantized operator is antiunitary. There are three possibilities, which are reviewed in detail in Refs. [9, 10].

The first one is *time-reversal symmetry*. Time reversal acts on the many-body Hamiltonian via an antiunitary operator $\hat{\mathcal{T}}$. In first-quantized language, the symmetry is realized via an antiunitary operator T that satisfies $THT^{-1} = H$, that is, $[H, T] = 0$. In the case of a Bloch Hamiltonian $H(k) = e^{-ik\cdot r}He^{ik\cdot r}$, time reversal gives $TH(k)T^{-1} = H(-k)$. Furthermore, there are two possibilities for the square of the operator T, namely $T^2 = \pm 1$. Summarizing, there are three ways in which a Hamiltonian may respond to time-reversal symmetry. These are classified by a number T that takes the following values:

$$\mathsf{T} = \begin{cases} 0, & \text{when the Hamiltonian is not time-reversal invariant,} \\ +1, & \text{when the Hamiltonian is time-reversal invariant and } T^2 = +1, \\ -1, & \text{when the Hamiltonian is time-reversal invariant and } T^2 = -1. \end{cases} \tag{1.11}$$

The second such symmetry is *particle-hole symmetry*. It acts on the many-body Hamiltonian through a *unitary* operator $\hat{\mathcal{C}}$, but in first-quantized language corresponds to an antiunitary operator C such that $CHC^{-1} = -H$, i.e., $\{H, C\} = 0$. For a Bloch Hamiltonian, particle-hole symmetry gives $CH(k)C^{-1} = -H(-k)$. Again, the operator C may square to $+1$ or -1, so we assign a number C that classifies the particle-hole character of the Hamiltonian:

$$\mathsf{C} = \begin{cases} 0, & \text{when the Hamiltonian is not particle-hole invariant,} \\ +1, & \text{when the Hamiltonian is particle-hole invariant and } C^2 = +1, \\ -1, & \text{when the Hamiltonian is particle-hole invariant and } C^2 = -1. \end{cases} \tag{1.12}$$

The combination of $\hat{\mathcal{T}}$ and $\hat{\mathcal{C}}$ may lead to a third symmetry, the so-called *chiral* or *sublattice symmetry*. The corresponding operator on the Fock space is defined as $\hat{\mathcal{S}} \equiv \hat{\mathcal{T}} \cdot \hat{\mathcal{C}}$, which is antiunitary. Chiral symmetry acts on the single-particle basis via a *unitary* operator $S \equiv T \cdot C$ that satisfies $SHS^{-1} = -H$, so again $\{H, S\} = 0$. This corresponds to $SH(k)S^{-1} = -H(-k)$

for a Bloch Hamiltonian. Note that, even though S acts on H in the same way as C does, the former is unitary whereas the latter is antiunitary. Unlike for the time-reversal and particle-hole symmetries, the chiral operator always satisfies $S^2 = +1$. The chiral character of H is given by a number S as follows:

$$
\mathsf{S} = \begin{cases} 0, & \text{when the Hamiltonian does not possess chiral symmetry,} \\ 1, & \text{when the Hamiltonian does possess chiral symmetry.} \end{cases} \tag{1.13}
$$

A system that breaks *both* time-reversal and particle-hole symmetry may still have chiral symmetry. However, when only *one* of T and C is broken, $\mathsf{S} = 0$ necessarily. In other words, whenever $\mathsf{T} \neq 0$ or $\mathsf{C} \neq 0$ the chiral character of the system is completely determined. However, when $\mathsf{T} = \mathsf{C} = 0$, we have both possibilities for S. Counting all possible combinations of T, C, and S gives a total of $(3 \times 3 - 1) + 2 = 10$ distinct *symmetry classes*. These are given in Table 1.1.

class	T	C	S
A	0	0	0
AIII	0	0	1
AI	+	0	0
BDI	+	+	1
D	0	+	0
DIII	−	+	1
AII	−	0	0
CII	−	+	1
C	0	−	0
CI	+	−	1

Table 1.1. The ten symmetry classes according to the different possible realizations of antiunitarily realized symmetries, namely: time reversal (T), particle-hole (C), and chiral symmetry (S). The plus and minus signs below T and C stand for the square of the corresponding operators T and C, respectively.

1.2.2 The tenfold way

We have established that a topological insulator is a state of matter which cannot be deformed to Dirac's vacuum without closing the energy gap. However, it is still unclear how many of these states there are, and how precisely they differ from the atomic insulator. Consider an insulating lattice system in d dimensions possessing some of the antiunitary symmetries described in the previous section, and *no* unitary symmetries[2]. The corresponding Hamiltonian will then unambiguously belong to a single symmetry class. Continuous deformations of this Hamiltonian will induce changes in the energy spectrum and the wave functions. However, these changes will be adiabatic *provided* we never close the energy gap and preserve the original antiunitary symmetries. We can now ask ourselves which Hamiltonians can be connected in this way. This question can be answered with tools from topology.

Before we begin, the problem can be simplified as follows. The Bloch Hamiltonian satisfies

$$
H(\mathbf{k})|u_a(\mathbf{k})\rangle = E_a(\mathbf{k})|u_a(\mathbf{k})\rangle, \tag{1.14}
$$

---

[2]In general, a system will possess some unitary symmetries such as rotational symmetry, parity symmetry, etc. However, here we assume that we have gotten rid of these by block-diagonalizing the Hamiltonian in subspaces of well-defined quantum numbers. We then focus on one of these individual blocks, whose remaining symmetries can only be antiunitary.

where a denotes the band index and $\boldsymbol{k}$ is the d-dimensional crystal momentum living in the first Brillouin zone, $\mathrm{BZ}^d \cong \mathbb{T}^d$. We consider the case of a ground state with n filled and m empty bands. Since we are interested in topological properties of the system, we may continuously deform the Hamiltonian into a simplified form where all filled bands have energy $E = -1$ and all empty bands have energy $E = +1$. We denote this simplified Hamiltonian by $\mathcal{Q}(\boldsymbol{k})$. All topological properties will remain unchanged by this continuous deformation.

To illustrate the role of topology in this problem we consider the simplest possible case, namely a system possessing no symmetry conditions at all. This corresponds to class A in Table 1.1. The simplified Hamiltonian can be written in the following form:

$$\mathcal{Q}(\boldsymbol{k}) = U(\boldsymbol{k}) \Lambda U(\boldsymbol{k})^\dagger, \quad \text{where} \quad \Lambda = \begin{bmatrix} \mathbb{I}_m & 0 \\ 0 & -\mathbb{I}_n \end{bmatrix}. \tag{1.15}$$

Here, $\mathbb{I}_n$ denotes the $n \times n$ identity matrix. The matrix Λ contains the eigenvalues and $U(\boldsymbol{k})^\dagger \in \mathrm{U}(m+n)$ is the unitary matrix of eigenvectors that diagonalizes $\mathcal{Q}(\boldsymbol{k})$. Consider a case where $U(\boldsymbol{k})$ is of the form

$$U(\boldsymbol{k}) = \begin{bmatrix} U_m(\boldsymbol{k}) & 0 \\ 0 & U_n(\boldsymbol{k}) \end{bmatrix}, \tag{1.16}$$

with $U_m(\boldsymbol{k}) \in \mathrm{U}(m)$ and $U_n(\boldsymbol{k}) \in \mathrm{U}(n)$. In this case, $\mathcal{Q}(\boldsymbol{k}) = \Lambda$ and the simplified Hamiltonian remains invariant with respect to the case when $U(\boldsymbol{k}) = \mathbb{I}_{m+n}$. There are no other instances in which $\mathcal{Q}(\boldsymbol{k})$ remains unchanged. Therefore, the matrix $U(\boldsymbol{k})$ can be regarded as an element of the coset space $\mathrm{U}(m+n)/(\mathrm{U}(m) \times \mathrm{U}(n))$. Hence, every ground state of $\mathcal{Q}(\boldsymbol{k})$ can be described by the map

$$\begin{aligned} U \colon \mathrm{BZ}^d &\to \mathrm{U}(m+n)/(\mathrm{U}(m) \times \mathrm{U}(n)) \\ \boldsymbol{k} &\mapsto U(\boldsymbol{k}). \end{aligned} \tag{1.17}$$

Every such map describes a ground state, that is, a Fermi sea of occupied states. We now ask ourselves how many distinct ground states of this kind there are, i.e., how many such maps exist that cannot be continuously deformed into each other. For the moment, let us assume for simplicity that the Brillouin zone is not a torus, but a d-dimensional sphere. In this case, we know from Sec. 1.1 that the answer is given by the homotopy group $\pi_d(\mathrm{U}(m+n)/(\mathrm{U}(m) \times \mathrm{U}(n)))$, which is known. For instance, when the spatial dimension is $d = 2$, one obtains $\pi_2(\mathrm{U}(m+n)/(\mathrm{U}(m) \times \mathrm{U}(n))) \cong \mathbb{Z}$. This means that for every integer there is a ground state, and ground states which are assigned different integers cannot be continuously deformed into each other without closing the energy gap of the bulk Hamiltonian $H(\boldsymbol{k})$. In three dimensions, one finds $\pi_3(\mathrm{U}(m+n)/(\mathrm{U}(m) \times \mathrm{U}(n))) \cong \{e\}$, so the homotopy group is trivial and all ground states are homotopically equivalent.

We have now seen how to classify the topologically inequivalent Hamiltonians in symmetry class A when the crystal momentum is taken to live on a sphere. In order to obtain a true physical classification, one ought to consider the Brillouin zone as a torus. The answer is then not given by a homotopy group, but rather by its generalization to maps from $\mathbb{T}^d$ to the appropriate classifying space. For the rest of the symmetry classes, the procedure follows in the same vein. The symmetries impose certain constraints on the form of the Hamiltonian, which in turn determine the corresponding classifying space. Once the correct classifying space has been found, topologically distinct ground states are identified via the homotopy classes of the maps from the Brillouin zone to this space. In this way, topological phases are viewed as the homotopy classes of continuous families of gapped Hamiltonians. This program has been carried out for all topological insulators and semiconductors of noninteracting fermions. Three entirely different approaches have been used, namely (i) Anderson localization by Schnyder,

Ryu, Furusaki, and Ludwig [11–13], (ii) K-theory by Kitaev [14], and (iii) quantum anomalies by Ryu, Moore, and Ludwig [15]. While the first two are only applicable to noninteracting systems, the classifying principle based on quantum anomalies extends to interacting theories. In this section we are mainly concerned with the noninteracting case. The resulting classification is shown in Table 1.2 and takes the form of a periodic table for topological insulators, commonly called the *tenfold way*. Given one of the ten symmetry classes and the dimensionality of space, one can readily identify whether this setting admits distinct topological phases and the kind of topological invariant that characterizes them. For $d \geq 8$, the table loops back starting from the first column, a phenomenon known as *Bott periodicity*.

Now that we have the tools to know which kinds of systems should host nontrivial topological phases, we are in position to briefly review two important physical examples of such states of matter. We begin with the quantum Hall effect in the following section and then proceed to discuss the quantum spin Hall effect.

class	symmetry			dimension							
	T	C	S	0	1	2	3	4	5	6	7
A	0	0	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0
AIII	0	0	1	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$
AI	+	0	0	$\mathbb{Z}$	0	0	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$
BDI	+	+	1	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$
D	0	+	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0	$2\mathbb{Z}$	0
DIII	−	+	1	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0	$2\mathbb{Z}$
AII	−	0	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0
CII	−	−	1	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0
C	0	−	0	0	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0
CI	+	−	1	0	0	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$

Table 1.2. Classification of topological insulators or "tenfold way". All gapped fermionic Hamiltonians are classified according to their antiunitary symmetries and the dimensionality of space. The symbols $\mathbb{Z}$, $2\mathbb{Z}$, and $\mathbb{Z}_2$ denote the kind of topological invariant that labels the distinct topological phases in a particular setting. A zero corresponds to the situation in which there are no nontrivial topological invariants, that is, all gapped Hamiltonians are homotopically equivalent to each other.

1.2.3 The quantum Hall effect

The simplest setup in which topology plays a crucial role in a condensed-matter system is that giving rise to the *integer quantum Hall effect* (IQHE), which we present in this section. The effect is extensively discussed in Refs. [16, Chs. 1–4, 17, Chs. 1–3]. We assume that the reader is familiar with the *classical* Hall effect, discovered in 1879 by Edwin Hall [18].

The IQHE can be observed in a setup such as that of Fig. 1.3 and was first discovered in 1980 by von Klitzing, Dorda, and Pepper [20, 21]. The setup is identical to that of its classical counterpart, with the difference that the IQHE is observed at low temperatures for strong values of the magnetic field B. We recall that the classical Drude model gives $\rho_{xx} = m/ne^2\tau$ and $\rho_{xy} = B/ne$ for the longitudinal and transverse conductivities, respectively. Here, τ is the scattering time (average time between electron collisions) and n is the electron density. While this holds for small magnetic fields, the IQHE is characterized by a deviation of this law for higher values of B. In the IQHE regime one finds that the transverse resistivity exhibits plateaux and the longitudinal resistivity vanishes whenever ρ_{xy} remains constant, as explained in the
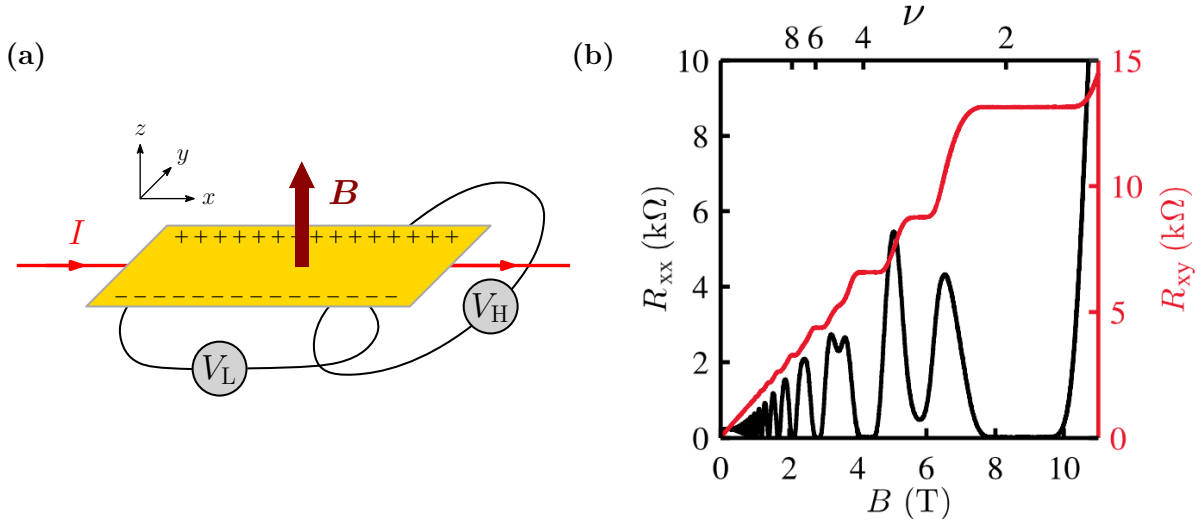
Figure 1.3. The integer quantum Hall effect. **(a)** Experimental setup for the observation of the IQHE. A semiconducting sample of length L and width W is crossed by an electric current of magnitude I in the x-direction. When turning on a magnetic field of magnitude B perpendicular to the sample, a measurable potential difference V_H appears in the y-direction. Classically, this is attributed to the fact that electrons deviate from their straight trajectories due to a nonvanishing Lorentz force, causing a net accumulation of charge at both sides of the sample. The associated transverse resistance, or Hall resistance, is $R_H = V_H/W$ and is equal to the transverse *resistivity* in this geometry. **(b)** Longitudinal and transverse resistances (R_{xx} and $R_{xy} = R_H$, respectively) as a function of the magnetic field. For small B, the longitudinal resistance remains constant and the transverse resistance grows linearly, which corresponds to a classical Hall effect. Increasing the magnetic field causes the Hall resistance to increase in discrete multiples of h/e^2 at certain values of B. More precisely, $R_H = \frac{h}{e^2}\frac{1}{\nu}$, where ν is an integer. The longitudinal resistivity shows a sharp peak at each of these sudden increases. In between these values of the magnetic field, R_H remains constant and the profile shows a series of plateaux with vanishing longitudinal resistivity. Source: Ref. [19].

caption of Fig. 1.3. At each of these plateaux, the transverse conductivity has the value

$$\rho_{xy} = \frac{h}{e^2}\frac{1}{\nu}, \quad \text{with} \quad \nu \in \mathbb{Z}. \tag{1.18}$$

Here, h is Planck's constant and $-e$ is the charge of the electron. In the limit $\tau \to \infty$, which is valid for pure enough samples, the conductivity is simply $\sigma_{xy} = \rho_{xy}^{-1}$. The quantization of σ_{xy} has been measured to one part in 10^9 [22].

It is now well known that the IQHE is a consequence of the quantum-mechanical behavior of a noninteracting electron gas in a magnetic field in combination with disorder arising from impurities, as originally shown by Laughlin [23]. The bulk of the sample mostly contains localized states bound to the impurities, while the edges are perfectly conducting thanks to the presence of extended states. These edge states are *chiral*: they propagate in only one direction along the boundary of the sample, either clockwise or counterclockwise. The corresponding dispersion relation is gapless and forms a conduit between the valence and conduction bands, as seen in Fig. 1.4 (a). While a detailed understanding of the physical origin of the IQHE is not necessary for our subsequent discussion, we want to draw attention to the following expression for the quantization of the conductivity, first derived by Thouless, Kohmoto, Nightingale, and den Nijs (TKNN) via the Kubo formula [24]:

$$\sigma_{xy} = \frac{e^2}{h}\sum_{n=1}^{N} \mathcal{C}_n. \tag{1.19}$$

The sum runs over all N filled bands, and for each of these $\mathcal{C}_n \in \mathbb{Z}$ is an integer called the

TKNN invariant or the (first) *Chern number*. It has the following expression in terms of the Bloch state $|u_{nk}\rangle$ of the nth occupied band[3]:

$$\mathcal{C}_n = -\frac{\mathrm{i}}{2\pi} \int_{\mathrm{BZ}} \mathrm{d}k_x \, \mathrm{d}k_y \left(\left\langle \frac{\partial u_{nk}}{\partial k_x} \middle| \frac{\partial u_{nk}}{\partial k_y} \right\rangle - \left\langle \frac{\partial u_{nk}}{\partial k_y} \middle| \frac{\partial u_{nk}}{\partial k_x} \right\rangle \right). \tag{1.20}$$

The key point is that the Chern number is actually a topological invariant. This explains the robustness of the quantum Hall state and the extremely accurate quantization of σ_{xy}. It must be noted that the Chern number of a single band is only defined when there are no accidental degeneracies with other bands, but the sum over all occupied bands is well-defined even in the presence of such degeneracies. An integer quantum Hall system belongs to class A in the classification of topological insulators of Table 1.2. Therefore, the above Chern number is precisely the topological invariant characterizing the different phases in this symmetry class in two dimensions. Similarly, the table immediately reveals that there should be no quantum Hall response in 3D, which is a well-known result.

In 1988, a toy model for a quantum Hall effect without magnetic field was devised by Haldane [27]. The Haldane model is defined on a honeycomb lattice and its crucial feature is a second-nearest-neighbor hopping that breaks time-reversal symmetry. In this way it was first established that the quantum Hall effect is not necessarily a consequence of an external magnetic field, but can rather be placed in a wider context of phenomena associated with broken time-reversal symmetry.

1.2.4 The quantum spin Hall effect

We now move on to the quantum *spin* Hall effect (QSHE), which was first introduced in 2005 in a seminal paper by Kane and Mele [28]. Their original model uses graphene, consisting of a honeycomb lattice of carbon atoms with two sublattices A and B. The Brillouin zone is hexagonal with the states near the Fermi energy residing near the K and K' points (or *valleys*) at opposite corners. The Hamiltonian describing the QSH system is given by

$$H(\boldsymbol{k}) = H_0(\boldsymbol{k}) + H_{\mathrm{SO}}(\boldsymbol{k}), \tag{1.21}$$

where

$$H_0(\boldsymbol{k}) = -\mathrm{i}v_{\mathrm{F}}(\sigma_x \tau_z k_x + \sigma_y k_y), \tag{1.22a}$$
$$H_{\mathrm{SO}}(\boldsymbol{k}) = \Delta_{\mathrm{SO}} \sigma_z \tau_z s_z. \tag{1.22b}$$

Here, the Pauli matrices σ_i act on the two sublattices of graphene, the τ_i act on the two valleys, and the s_i act on the spin degree of freedom. More precisely, $\sigma_z = +1 \, (-1)$ describes states on the A (B) sublattice, $\tau_z = +1 \, (-1)$ describes states at the K (K') point, and $s_z = +1$ (-1) describes states with up (down) spin. Note that Kronecker products are understood everywhere. The term $H_0(\boldsymbol{k})$ describes the effective-mass approximation valid near the two valleys, whereas $H_{\mathrm{SO}}(\boldsymbol{k})$ describes the spin-orbit interaction. The parameter v_{F} is called the *Fermi velocity*, since it gives the group velocity of the gapless states described by $H_0(\boldsymbol{k})$ with dispersion $\epsilon_0(\boldsymbol{k}) = \pm v_{\mathrm{F}}|\boldsymbol{k}|$. The spin-orbit term leads to a gap $2|\Delta_{\mathrm{SO}}|$, and the dispersions of the total Hamiltonian are given by $\epsilon(\boldsymbol{k}) = \pm\sqrt{v_{\mathrm{F}}^2|\boldsymbol{k}|^2 + \Delta_{\mathrm{SO}}^2}$.

---

[3]Technically speaking, generators of translations do not commute with one another in the presence of a magnetic field and one cannot directly apply Bloch's theorem. However, one can define a "magnetic lattice" where each unit cell has area hc/eB and encloses a single flux quantum, in such a way that lattice translations commute. Bloch's theorem is then restored and one can write the eigenstates in terms of *magnetic* Bloch functions with crystal momentum $\boldsymbol{k}$. More information on this point can be found in Refs. [25, 26].

It is straightforward to verify that the following operators implement the time-reversal and particle-hole symmetries, respectively:

$$T = i\tau_x s_y \mathcal{K}, \quad \text{with} \quad T^2 = -1, \tag{1.23a}$$

$$C = \tau_x \sigma_z \mathcal{K}, \quad \text{with} \quad C^2 = +1. \tag{1.23b}$$

Here, $\mathcal{K}$ denotes the complex conjugation operator, that is, $\mathcal{K}i\mathcal{K}^{-1} = -i$ and $\mathcal{K}^2 = 1$. If we do not allow for perturbations that break these symmetries, the model has $T = -1$ and $C = +1$, and thus $S = 1$. It therefore belongs to the symmetry class DIII in Table 1.2. The corresponding topological invariant in 2D is denoted by ν and is either 0 or 1, i.e., $\nu \in \mathbb{Z}_2$. In addition to these antiunitary symmetries, the model also has inversion symmetry. Because of this, the topological invariant can be determined as the product of parities of the Bloch functions of all occupied bands at the four time-reversal-invariant momenta (TRIM) [29, 30]:

$$(-1)^\nu = \prod_{i=1}^{4} \delta_i, \quad \text{where} \quad \delta_i = \prod_{m=1}^{N} \xi_{2m}(\Gamma_i). \tag{1.24}$$

Here, $\xi_{2m}(\Gamma_i)$ is the parity eigenvalue of the $2m$th occupied band at the TRIM Γ_i. It shares the same eigenvalue with its degenerate time-reversed partner arising from the famous Kramers theorem, which states that every energy level of a time-reversal-symmetric system with half-integer spin is at least doubly degenerate [31, 32]. The above expression for ν can be cast into the following form involving the *Pfaffian* of a matrix $\mathcal{W}$:

$$(-1)^\nu = \prod_{i=1}^{4} \frac{\text{Pf}[\mathcal{W}(\Gamma_i)]}{\sqrt{\det[\mathcal{W}(\Gamma_i)]}}, \quad \text{with} \quad \mathcal{W}_{\alpha\beta}(\mathbf{k}) = \langle u_\alpha(-\mathbf{k})|T|u_\beta(\mathbf{k})\rangle. \tag{1.25}$$

The system presented by Kane and Mele once again hosts edge states. Unlike the chiral edge states of the QHE, however, the QSHE is characterized by the appearance of *helical* edge states. In this scenario there are states propagating in both directions at each edge, and these are "spin-filtered", that is, electrons with opposite spin propagate in opposite directions. This phenomenon is known by the name of *spin-momentum locking* and is illustrated in Fig. 1.4 (b). The topological invariant ν counts the parity of the number of Kramers pairs of states at each edge [33, 34]. A nontrivial QSH phase is characterized by an odd number of such pairs.

It is worth mentioning that the parameter Δ_{SO} in graphene is far too small to actually observe a QSH phase. However, almost immediately after the publication of Kane and Mele's paper, Bernevig, Hughes, and Zhang showed that a QSH state could be realized in HgTe–CdTe quantum wells [35]. A detailed review on theoretical and experimental aspects of the QSHE can be found in Ref. [36].

1.2.5 The bulk-boundary correspondence

The existence of edge states at the boundary of a topological insulator follows from to the so-called *bulk-boundary correspondence*. While we will not go into much detail, this fundamental principle relates a nontrivial topological property of the bulk with a physical observable at the boundary. For instance, the absolute value of the Chern number in the QHE gives the number of edge states that propagate along the edges of the sample. Physically, it is easy to realize that *something* special must happen at the boundary between two topologically distinct phases. The different topological invariants force the wave function to acquire a "twist" when crossing the boundary between the two samples. The physical realization of this twist turns out to come in the form of localized edge states. More information on the bulk-boundary correspondence can be found in Refs. [37, 38].
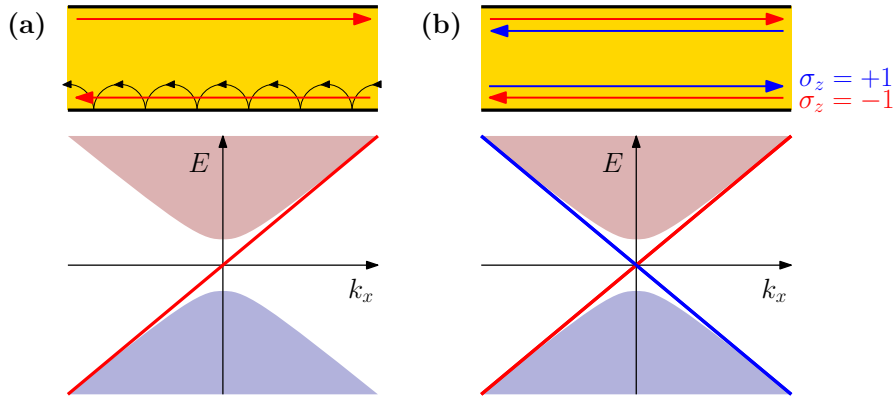
Figure 1.4. Comparison of the quantum Hall and quantum spin Hall effects. **(a)** In the QHE, chiral electronic edge states move along the edges of the sample. This motion can be classically understood in the sense of "skipping orbits", depicted in black: the Lorentz force causes electrons to move in circles, but those that are close to the edges cannot complete a full circular motion and skip forward instead. The band diagram shows the dispersion relation of the chiral state at the top edge of the sample, which crosses the gap between the bulk valence and conduction bands. **(b)** In the QSHE, there is no net charge transport along the edges. The edge states are not chiral, but helical: up- and down-spin electrons (depicted in blue and red, respectively) move in opposite directions along the edge, producing a net spin transport. Again, the band diagram shows the dispersion of the helical states propagating along the top edge of the sample. Both bands cross the energy gap in opposite directions, in accord with the opposite group velocities of the corresponding electrons.

1.2.6 The role of interactions

So far, our exposition of topological insulators has not included many-body interactions of any kind. In this simplified case, the classification of fermionic topological phases is very well understood. However, a lot can be asked about topological phenomena in interacting systems. First and foremost, it is natural to wonder if the free-fermion topological phases are stable under the inclusion of interactions. Another interesting question is if there are generalizations of topological insulators without a free-fermion analog. More generally, one is led to ask how the free-fermion classification of Table 1.2 is altered in strongly correlated systems.

Noninteracting topological insulators and superconductors are commonly studied via the formalism of *topological band theory* [33, 39], which is not suited for the analysis of interacting systems. For this reason, Qi, Hughes, and Zhang first developed a *topological field theory* of time-reversal invariant insulators, which is also apt for the study of interactions [40]. Another powerful approach to the classification of correlated topological phases is based on *quantum anomalies*, the phenomenon by which a classical global or gauge symmetry is lost in the quantization process [15, 41–43]. It is now known, for instance, that in 3D the classification of class DIII collapses from $\mathbb{Z}$ to $\mathbb{Z}_{16}$ when interactions are added, and similar collapses have been found in classes AIII and CI in three dimensions [44]. These reductions to smaller groups make sense if we take into account that adding interactions will in general influence the energy gap. In this way, two phases that were previously topologically inequivalent may now be connected without closing the gap, which is kept finite thanks to interaction effects.

In particular, the stability of a quantum spin Hall phase in the presence of interactions has been addressed in Refs. [45–47]. These articles consider the effect of weak disorder and interactions on the edge modes of a QSH state. All studies agree that in a QSH system with a nontrivial $\mathbb{Z}_2$ topological invariant, the edge states are robust against these perturbations. However, the topological distinction between phases with an even or odd number of Kramers pairs seems to disappear in the presence of strong interactions.

1.3 Berry physics

The Chern number we have introduced in Eq. (1.20) is rooted in the mathematical theory of fiber bundles [3]. However, it can be understood physically in terms of the *Berry phase* [48]. An in-depth review of the topic can be found in Refs. [2, Ch. 6, 49, 50, Ch. 3]; here we outline the main ideas.

1.3.1 Adiabatic evolution of a quantum system

Consider a physical system described by a Hamiltonian that depends on time through a set of parameters $\boldsymbol{R} = (R_1, R_2, \dots)$, that is, $H(t) = H(\boldsymbol{R}(t))$. We let the system evolve adiabatically in time as $\boldsymbol{R}(t)$ slowly moves along a path $\mathcal{P}$ through parameter space. It is useful to introduce an instantaneous orthonormal basis of eigenstates $|u_n(\boldsymbol{R})\rangle$, that is, $H(\boldsymbol{R})|u_n(\boldsymbol{R})\rangle = \epsilon_n(\boldsymbol{R})|u_n(\boldsymbol{R})\rangle$. Note that this alone does not completely determine the basis functions, as we can always make a *gauge transformation* by multiplying $|u_n(\boldsymbol{R})\rangle$ by some $\boldsymbol{R}$-dependent phase. We require that this phase is smooth and single-valued along $\mathcal{P}$.

According to the quantum adiabatic theorem [51], a system initially in one of its eigenstates $|u_n(\boldsymbol{R}(0))\rangle$ will remain as an eigenstate of the Hamiltonian $H(\boldsymbol{R}(t))$ throughout the evolution. This holds when the eigenstate in question does not become degenerate with another eigenstate at some point of the process, which will be assumed to be the case. The only degree of freedom is therefore the phase of the quantum state. We write the state at time t as

$$|\psi_n(t)\rangle = \mathrm{e}^{-\mathrm{i}\gamma_n(t)} \exp\left[-\mathrm{i}\int_0^t \mathrm{d}t'\, \epsilon_n(\boldsymbol{R}(t'))\right]|u_n(\boldsymbol{R}(t))\rangle\,. \tag{1.26}$$

The second exponential is just the dynamical phase factor, but there may be an additional contribution $\gamma_n(t)$. Inserting this expression for $|\psi_n(t)\rangle$ into the time-dependent Schrödinger equation, one finds that this extra phase can be expressed as a path integral in parameter space as follows:

$$\gamma_n = \int_{\mathcal{P}} \mathrm{d}\boldsymbol{R} \cdot \boldsymbol{\mathcal{A}}_n(\boldsymbol{R})\,. \tag{1.27}$$

Here, $\boldsymbol{\mathcal{A}}_n(\boldsymbol{R})$ is the so-called the *Berry connection*, which is the vector-valued function

$$\boldsymbol{\mathcal{A}}_n(\boldsymbol{R}) = -\mathrm{i}\langle u_n(\boldsymbol{R})|\boldsymbol{\nabla}_{\boldsymbol{R}}|u_n(\boldsymbol{R})\rangle\,. \tag{1.28}$$

The Berry connection is gauge-dependent. Indeed, under a gauge transformation

$$|u_n(\boldsymbol{R})\rangle \to \mathrm{e}^{\mathrm{i}\zeta(\boldsymbol{R})}|u_n(\boldsymbol{R})\rangle\,, \tag{1.29}$$

with $\zeta(\boldsymbol{R})$ an arbitrary smooth function, $\boldsymbol{\mathcal{A}}_n(\boldsymbol{R})$ transforms as

$$\boldsymbol{\mathcal{A}}_n(\boldsymbol{R}) \to \boldsymbol{\mathcal{A}}_n(\boldsymbol{R}) + \boldsymbol{\nabla}_{\boldsymbol{R}}\zeta(\boldsymbol{R})\,. \tag{1.30}$$

Let us now consider a *cyclic* evolution, with the system returning to the original parameter-space point at time T, that is, $\boldsymbol{R}(T) = \boldsymbol{R}(0)$. The requirement that $\mathrm{e}^{\mathrm{i}\zeta(\boldsymbol{R})}$ be single-valued implies that any gauge transformation must be subject to the constraint $\zeta(\boldsymbol{R}(T)) - \zeta(\boldsymbol{R}(0)) \in 2\pi\mathbb{Z}$. It follows that the phase γ_n can only be changed by an integer multiple of 2π and can in general not be fully removed. Thus, for a closed path, γ_n becomes a quantity which is gauge-invariant modulo 2π. This quantity is known as the *Berry phase*:

$$\gamma_n = \oint_{\mathcal{P}} \mathrm{d}\boldsymbol{R} \cdot \boldsymbol{\mathcal{A}}_n(\boldsymbol{R})\,. \tag{1.31}$$

It is clear from this definition that the Berry phase only depends on the geometric aspect of the closed path. It is independent of how $\boldsymbol{R}(t)$ varies in time as long as the evolution is slow enough for the adiabatic theorem to apply. Hence, the explicit time dependence is not essential in our subsequent discussion and will be dropped henceforth.

1.3.2 The Berry curvature

In analogy with electrodynamics, one can define a gauge-field tensor derived from the Berry connection:

$$\Omega^n_{\mu\nu}(\boldsymbol{R}) = \frac{\partial}{\partial R^\mu}\mathcal{A}^n_\nu(\boldsymbol{R}) - \frac{\partial}{\partial R^\nu}\mathcal{A}^n_\mu(\boldsymbol{R}) = -\mathrm{i}\left(\left\langle\frac{\partial u_n}{\partial R_\mu}\bigg|\frac{\partial u_n}{\partial R_\nu}\right\rangle - \left\langle\frac{\partial u_n}{\partial R_\nu}\bigg|\frac{\partial u_n}{\partial R_\mu}\right\rangle\right). \tag{1.32}$$

This field is called the *Berry curvature*, and it is manifestly gauge-invariant under (1.30). If the parameter space can be regarded as three-dimensional, one can represent the Berry curvature as a three-vector $\boldsymbol{\Omega}_n(\boldsymbol{R}) = \boldsymbol{\nabla_R} \times \boldsymbol{\mathcal{A}}_n(\boldsymbol{R})$ such that $\Omega^n_{ij} = \epsilon_{ijk}\Omega^n_k$.

Using the generalized Stokes theorem, the Berry phase (1.31) can now be written as the surface integral

$$\gamma_n = \frac{1}{2}\int_{\mathcal{S}} \mathrm{d}R^\mu \wedge \mathrm{d}R^\nu\, \Omega^n_{\mu\nu}(\boldsymbol{R}), \tag{1.33}$$

where $\mathcal{S}$ is an arbitrary surface bounded by the path $\mathcal{P}$. Before moving on, let us comment on a potentially puzzling aspect of gauge invariance regarding Eqs. (1.31) and (1.33). As we have said, the first of these equations is only gauge-invariant modulo 2π. However, the Berry curvature is *fully* gauge-invariant, and thus so is the Berry phase as expressed in Eq. (1.33). We are forced to ask whether γ_n is fully gauge-invariant or not. The answer is that, if γ_n is to be determined with knowledge of $|u_n(\boldsymbol{R})\rangle$ *only on the curve* $\mathcal{P}$, then it is only well-defined modulo 2π. The Berry phase determined from Eq. (1.31) will then agree with that of Eq. (1.33) only for a particular choice of gauge, while other choices will leave a mismatch of $2\pi m$, with m an integer. The question now is which gauge choice is "correct". The solution is to choose a gauge transformation that is smooth and continuous *everywhere in $\mathcal{S}$* (including its boundary $\mathcal{P}$). In this case both equations will agree. By contrast, a gauge that shifts γ_n by an integer multiple of 2π when regarding $|u_n(\boldsymbol{R})\rangle$ as a function defined only in the neighborhood of $\mathcal{P}$ cannot be smoothly continued into all the interior of $\mathcal{S}$. Indeed, in such a case one is forced to introduce vortex-like singularities which technically do not allow for the application of Stokes' theorem to go from (1.31) to (1.33).

Note that so far we have only discussed the case where a single energy level can be separated from the rest at all points of the adiabatic evolution. However, when there exist accidental degeneracies between levels, the dynamics must be projected to the subspace spanned by the corresponding eigenstates. In this situation a non-Abelian Berry curvature arises [52]. We will not need to be concerned with this issue in the rest of this work.

1.3.3 The Chern theorem

We have seen that the Berry phase is fully gauge-invariant when computed as the integral of the Berry curvature over the manifold $\mathcal{S}$. It turns out that this phase is quantized in integer multiples of 2π; this result is the famous *Chern theorem*. The corresponding integer, which we denote by $\mathcal{C}_n$ for the eigenstate $|u_n(\boldsymbol{R})\rangle$, is called the *Chern number* or *Chern index*. It is a topological invariant attached to the manifold of states $|u_n(\boldsymbol{R})\rangle$ defined over the surface $\mathcal{S}$. Therefore, manifolds with different Chern numbers are homotopically inequivalent.

In this thesis we will mostly be concerned with two-dimensional materials. A natural candidate for the parameter space in this case is therefore the two-dimensional Brillouin zone of crystal momenta $\boldsymbol{k}$. Knowledge of the band structure of the system under consideration allows us to compute the Chern number of individual bands, which reads

$$\mathcal{C}_n = \frac{1}{2\pi}\int_{\mathrm{BZ}} \mathrm{d}^2k\, \Omega^n_{xy}(\boldsymbol{k}). \tag{1.34}$$

Note that this is precisely the result of Eq. (1.20) for the quantization of the Hall conductivity. A Bloch state $|u_n(k)\rangle$ with a given Chern number cannot be topologically equivalent to a Bloch state with a different Chern number. In particular, Bloch states with a nonzero Chern number cannot be equivalent to a state in the Dirac vacuum. Hence, the Chern number provides us with a way to probe whether or not a given Hamiltonian describes a topological insulator.

1.4 Bi₂Se₃ as a topological insulator

In this section we introduce bismuth selenide (Bi₂Se₃), a transition metal chalcogenide that turns out to be a topological insulator. We closely follow Refs. [53, 54] to provide an overview of the physical phenomena that ultimately lead to the nontrivial bulk topology of this material.

1.4.1 Crystal structure, symmetries, and energy bands

The topological properties of Bi₂Se₃ can be understood from the interactions between the outermost electrons at the atomic sites of the crystal. To see which ones are most relevant, one must take into account the crystalline structure of the material.
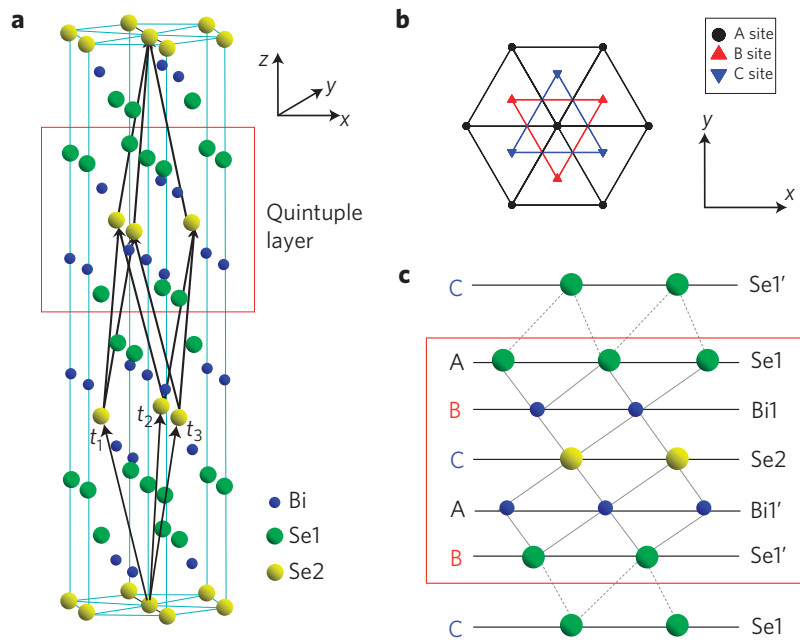

Figure 1.5. (a) Crystal structure of Bi₂Se₃, with primitive lattice vectors t_1, t_2, and t_3. A rhombohedral unit cell is shown, delimited by black arrows. **(b)** Projection of the triangle lattice on the xy-plane, with three possible positions A, B, and C. **(c)** Side view of the quintuple-layer structure. The stacking order of Se and Bi layers along the z-direction is $\cdots$–C(Se1′)–A(Se1)–B(Bi1)–C(Se2)–A(Bi1′)–B(Se1′)–C(Se1)–$\cdots$. Source: Ref. [53].

The crystal structure of Bi₂Se₃ is rhombohedral. As shown in Fig. 1.5 (a), the crystal has a layered structure stacked along the z-direction, and one unit cell contains two bismuth (Bi) atoms and three selenium (Se) atoms. The crystal has inversion symmetry along the z-axis, which makes the Bi atoms equivalent to each other, and similarly for two of the Se atoms. We denote these by Bi1 and Bi1′, and by Se1 and Se1′, respectively. The remaining Se atom, which we call Se2, is located between the two Bi atoms and is inequivalent to Se1 and Se1′. Five atomic layers can be viewed as one unit, which is usually called a *quintuple layer* (QL), and consists of the stack Se1–Bi1–Se2–Bi1′–Se1′. Each atomic layer forms a triangle lattice with three possible

positions with respect to the z-axis, denoted as A, B, and C, as shown in Fig. 1.5 (b). These layers are stacked in the order $\cdots$–A–B–C–A–B–C–$\cdots$ along the z-direction. Hence, if in a given layer the A sites are occupied by Se2 atoms, then in the following layer the Se2 atoms will occupy either the B or C sites, but not the A sites. This implies that none of the primitive vectors is directed along the z-axis. Our coordinate system is shown in Fig. 1.5 (c), and the primitive vectors are given by

$$t_1 = \left(0, \sqrt{3}a/3, c/3\right), \tag{1.35a}$$

$$t_2 = \left(-a/2, -\sqrt{3}a/6, c/3\right), \tag{1.35b}$$

$$t_3 = \left(a/2, -\sqrt{3}a/6, c/3\right). \tag{1.35c}$$

Here, a is the lattice constant in the xy-plane and c is the lattice constant along the z-direction.

The Bi$_2$Se$_3$ crystal has the following spatial symmetries:

(1) Three-fold rotation symmetry around the z-axis, R_3. It is given by the transformation $(x, y, z) \mapsto (x \cos\theta - y \sin\theta, x \sin\theta + y \cos\theta, z)$, with $\theta = 2\pi/3$, which sends $t_1 \mapsto t_2$, $t_2 \mapsto t_3$ and $t_3 \mapsto t_1$.

(2) Two-fold rotation along the x-direction, R_2. It is given by the transformation $(x, y, z) \mapsto (x, -y, -z)$ along with the exchanges Bi1 $\leftrightarrow$ Bi1$'$ and Se1 $\leftrightarrow$ Se1$'$.

(3) Inversion, P. It is given by the transformation $(x, y, z) \mapsto (-x, -y, -z)$ along with the exchanges Bi1 $\leftrightarrow$ Bi1$'$ and Se1 $\leftrightarrow$ Se1$'$. The Se2 site plays the role of an inversion center of this lattice structure, so it is customary to choose it as the origin point.

(4) Time-reversal, T. This can be seen from the fact that there exists no spontaneous magnetization of the atoms and hence no intrinsic magnetic field on the lattice.

Ab initio calculations of the band structure of Bi$_2$Se$_3$ as performed in Refs. [53, 54] show that the inclusion of spin-orbit coupling (SOC) produces a *band inversion* at the Γ point. This is a phenomenon by which the conduction and valence bands exchange their parities due to an avoided crossing. Fig. 1.6 shows the band structure with and without SOC. Even though band inversion is usually a necessary mechanism to obtain topological materials, topological non-triviality can only be firmly established by the identification of a nonzero topological invariant. We have seen that, in time-reversal-invariant insulators, the product of parities of the Bloch wave functions of the occupied bands at all time-reversal-invariant momenta in the Brillouin zone is a topological invariant. In Bi$_2$Se$_3$, the parity of one band at the Γ point changes when turning on the SOC, whereas it remains unchanged for all occupied bands at the other TRIM points. Since the system without SOC, and hence without band inversion, is guaranteed to be a trivial insulator, Bi$_2$Se$_3$ is identified as a topological insulator.

1.4.2 Physics behind the band inversion

Keeping in mind the crystalline structure of Bi$_2$Se$_3$, we now proceed to analyze the interaction effects between the outer-shell electrons of the Bi and Se atoms. This will allow us to understand the band structure of the material as well as the origin of its topological properties. Indeed, by systematically including all relevant couplings, the band inversion described in Sec. 1.4.1 naturally appears. A sketch of the different steps is shown in Fig. 1.7.

We start from the atomic orbitals of Bi and Se and consider a tight-binding model for the corresponding valence electrons around the Γ point. The electron configuration of Bi is $6s^2 6p^3$, and that of Se is $4s^2 4p^4$. Since the outermost shells of both types of atoms are p orbitals, it is natural to only consider these p orbitals and neglect the rest. Due to the layered structure of Bi$_2$Se$_3$, the chemical bonding between atoms within one QL is very strong, whereas the
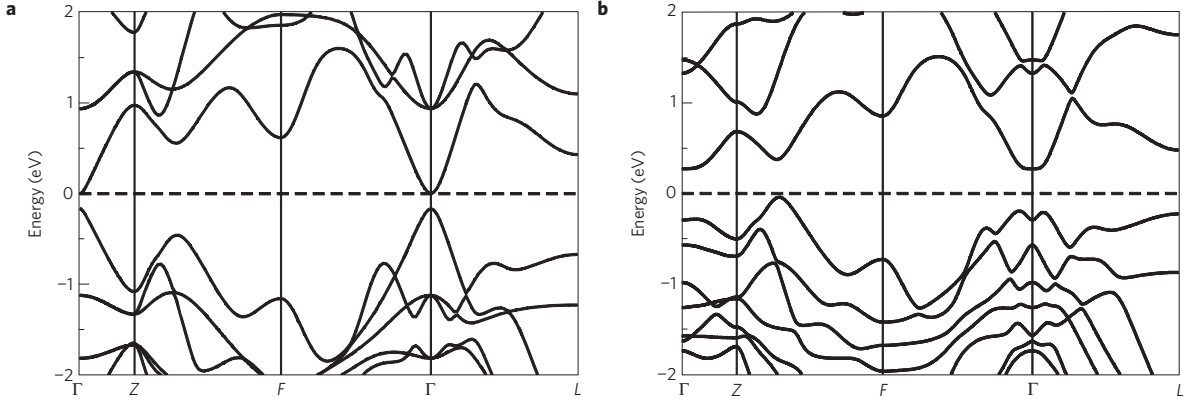
Figure 1.6. Band structure of Bi$_2$Se$_3$, **(a)** without spin-orbit coupling, and **(b)** with spin-orbit coupling. While most qualitative features remain unchanged when switching on the SOC, a very clear difference is observed at the Γ point, where the lowest conduction band and the highest valence band become inverted as a result of an avoided crossing. Source: Ref. [53].
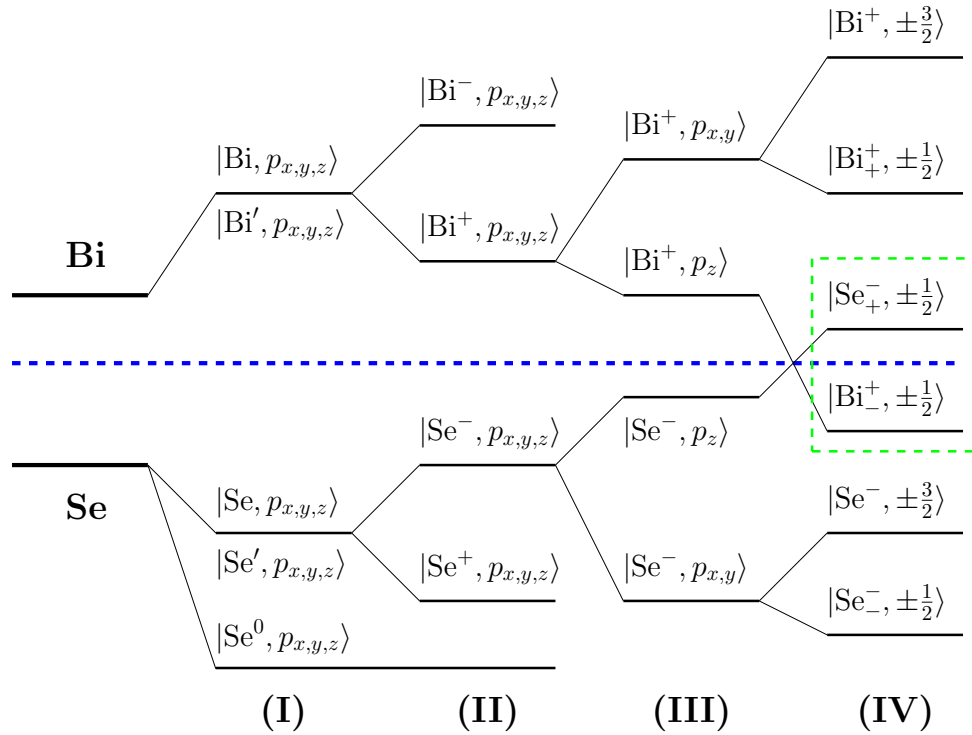

Figure 1.7. Schematic diagram of the evolution from the atomic $p_{x,y,z}$ orbitals of Bi and Se into the conduction and valence bands of Bi$_2$Se$_3$ at the Γ point. The blue dashed line represents the Fermi level. The following four steps are necessary to understand the band structure: **(I)** Hybridization of Bi orbitals and Se orbitals. **(II)** Formation of bonding and antibonding states due to parity symmetry. **(III)** Crystal field splitting arising from the asymmetry between the z- and x,y-directions. **(IV)** Influence of the SOC, which, if strong enough, may induce a level crossing leading to band inversion. Figure adapted from Ref. [54].

coupling between two neighboring QLs is predominantly due to van der Waals forces and hence relatively weak in comparison. Therefore, we first focus on one QL, which has five atoms in one unit cell. Each atom has three orbitals, namely p_x, p_y, and p_z. We first neglect the spin of the electrons in these orbitals; we will include it later in the discussion when we incorporate

the effect of SOC. Hence we denote the different orbitals by

$$|\Lambda, \alpha\rangle, \quad \text{where} \quad \begin{cases} \Lambda \in \{\text{Se1, Bi1, Se2, Bi1}', \text{Se1}'\}, \\ \alpha \in \{p_x, p_y, p_z\}. \end{cases} \tag{1.36}$$

As seen in the previous section, the Se2 atoms are located in the middle of each QL and are sandwiched between the two Bi layers, while the remaining Se atoms sit on the two outermost layers. Since all the Se layers are separated by Bi layers, the strongest coupling in this system stems from the interaction between Bi and Se layers. Without interaction, there are so far two energy levels in the system, namely those of the p-shell electrons in each type of atom. Including the aforementioned Bi–Se coupling causes level repulsion and breaks the degeneracy of the Se levels, because Se1 and Se1$'$ are inequivalent to Se2. Consequently, the Bi levels are pushed up and new hybridized states $|\text{Bi}, \alpha\rangle$ and $|\text{Bi}', \alpha\rangle$ ($\alpha \in \{p_x, p_y, p_z\}$) of equal energy are formed. On the other hand, the Se levels are pushed down. The Se1 and Se1$'$ orbitals hybridize into states $|\text{Se}, \alpha\rangle$ and $|\text{Se}', \alpha\rangle$ with equal energies, while the Se2 orbitals hybridize into states $|\text{Se}^0, \alpha\rangle$. These have lower energy than $|\text{Se}, \alpha\rangle$ and $|\text{Se}', \alpha\rangle$ because the sandwiching of the Se2 atoms between the two Bi layers causes a stronger level repulsion. The new states are shown in Fig. 1.7 (I).

The inversion symmetry of the system makes it convenient to combine the hybridized orbitals into bonding and antibonding states with definite parity. These read

$$|\text{Bi}^\pm, \alpha\rangle = \frac{1}{\sqrt{2}} \left(|\text{Bi}, \alpha\rangle \mp |\text{Bi}', \alpha\rangle \right), \tag{1.37a}$$

$$|\text{Se}^\pm, \alpha\rangle = \frac{1}{\sqrt{2}} \left(|\text{Se}, \alpha\rangle \mp |\text{Se}', \alpha\rangle \right), \tag{1.37b}$$

with the upper index denoting the parity eigenvalue. (Note that the p orbitals are odd under inversion, and that the parity operator also exchanges the primed and unprimed atoms.) These bonding and antibonding states are split upon taking into account the coupling between $|\text{Bi}, \alpha\rangle$ ($|\text{Se}, \alpha\rangle$) and $|\text{Bi}', \alpha\rangle$ ($|\text{Se}', \alpha\rangle$), with the antibonding state having higher energy than the bonding state in both cases. As shown in Fig. 1.7 (II), this implies that the states $|\text{Bi}^+, \alpha\rangle$ and $|\text{Se}^-, \alpha\rangle$ are closest to the Fermi surface, so we henceforth focus solely on these six states.

Next, we must take into account the fact that the layered structure makes the z-direction fundamentally different from the x- and y-directions in the atomic plane. This causes a further energy splitting between the p_z and the $p_{x,y}$ orbitals for both Bi$^+$ and Se$^-$ states, a phenomenon known as *crystal field splitting*. *Ab initio* calculations show that the energy of the $|\text{Bi}^+, p_{x,y}\rangle$ states is higher than that of the $|\text{Bi}^+, p_z\rangle$ states, whereas the energy of $|\text{Se}^-, p_{x,y}\rangle$ is lower than that of $|\text{Se}^-, p_z\rangle$. In this scenario, the conduction band would consist mostly of electrons in $|\text{Bi}^+, p_z\rangle$ orbitals, while the valence band would be dominated by $|\text{Se}^-, p_z\rangle$. This can be seen in Fig. 1.7 (III).

The last crucial missing ingredient is the spin-orbit coupling. As mentioned at the outset, we now introduce an additional spin index $\sigma \in \{\uparrow, \downarrow\}$ and label the states by $|\text{Bi}^+, \alpha, \sigma\rangle$ and $|\text{Se}^-, \alpha, \sigma\rangle$. The SOC Hamiltonian is given by

$$H_{\text{so}} = \lambda(r) \mathbf{L} \cdot \mathbf{S}, \tag{1.38}$$

where $\lambda(r) = \frac{1}{2m_e^2 c^2} \frac{1}{r} \frac{\partial U}{\partial r}$ depends on the detailed potential U between the outer-shell electrons and the rest of the atom. It is convenient to transform the p_x and p_y orbitals into $p_\pm$ orbitals with definite orbital angular momentum. These are given by

$$|\Lambda, p_\pm, \sigma\rangle = \frac{1}{\sqrt{2}} \left(|\Lambda, p_x, \sigma\rangle \pm i\, |\Lambda, p_y, \sigma\rangle \right), \quad \text{where now} \quad \Lambda \in \{\text{Bi}^+, \text{Se}^-\}. \tag{1.39}$$

These satisfy $L_z |\Lambda, p_\pm, \sigma\rangle = \pm |\Lambda, p_\pm, \sigma\rangle$. With the help of the Clebsch-Gordan tables we can find the matrix elements of the SOC Hamiltonian in this basis. Note that H_{so} commutes with J_z, the third component of the total angular momentum $\mathbf{J} = \mathbf{L} + \mathbf{S}$, and hence the angular momentum along the z-direction is still conserved. In other words, the matrix elements of H_{so} between subspaces of different eigenvalues of J_z all vanish. The remaining, possibly nonzero matrix elements read

$$\langle \Lambda, p_+, \uparrow | H_{\text{so}} |\Lambda, p_+, \uparrow\rangle = \langle \Lambda, p_-, \downarrow | H_{\text{so}} |\Lambda, p_-, \downarrow\rangle = \lambda_\Lambda/2, \tag{1.40a}$$

$$\langle \Lambda, p_+, \downarrow | H_{\text{so}} |\Lambda, p_+, \downarrow\rangle = \langle \Lambda, p_-, \uparrow | H_{\text{so}} |\Lambda, p_-, \uparrow\rangle = -\lambda_\Lambda/2, \tag{1.40b}$$

$$\langle \Lambda, p_+, \downarrow | H_{\text{so}} |\Lambda, p_z, \uparrow\rangle = \langle \Lambda, p_-, \uparrow | H_{\text{so}} |\Lambda, p_z, \downarrow\rangle = \lambda_\Lambda/\sqrt{2}, \tag{1.40c}$$

$$\langle \Lambda, p_z, \downarrow | H_{\text{so}} |\Lambda, p_z, \downarrow\rangle = \langle \Lambda, p_z, \uparrow | H_{\text{so}} |\Lambda, p_z, \uparrow\rangle = 0. \tag{1.40d}$$

Here, λ_Λ is a linear combination of the SOC coefficients for Bi and Se which depends on how much the respective orbitals are mixed into the state Λ, but is always positive because the effective potential U is necessarily attractive. At this point, a further hybridization takes place, but the conservation of J_z greatly constrains which orbitals can mix together. Specifically, $|\Lambda, +\frac{3}{2}\rangle = |\Lambda, p_+, \uparrow\rangle$ and $|\Lambda, -\frac{3}{2}\rangle = |\Lambda, p_-, \downarrow\rangle$ are still eigenstates of the Hamiltonian, whereas $|\Lambda, p_z, \uparrow\rangle$ ($|\Lambda, p_z, \downarrow\rangle$) mixes with $|\Lambda, p_+, \downarrow\rangle$ ($|\Lambda, p_-, \uparrow\rangle$) to give new hybridized orbitals, whose corresponding states we denote by $|\Lambda_\pm, +\frac{1}{2}\rangle$ ($|\Lambda_\pm, -\frac{1}{2}\rangle$). If we denote the energies after step (III) (right before the incorporation of the SOC) by E_x^Λ (degenerate with E_y^Λ) and E_z^Λ, then introducing H_{so} shifts E_x^Λ upwards by $\lambda_\Lambda/2$, whereas the other eigenenergies are obtained by diagonalizing the following 2×2 Hamiltonian:

$$H_\Lambda = \begin{bmatrix} E_x^\Lambda - \lambda_\Lambda/2 & \lambda_\Lambda/\sqrt{2} \\ \lambda_\Lambda/\sqrt{2} & E_z^\Lambda \end{bmatrix}. \tag{1.41}$$

In the end, the new energy levels read

$$E_{3/2}^\Lambda = E_x^\Lambda + \lambda_\Lambda/2, \tag{1.42a}$$

$$E_{1/2}^{\Lambda_\pm} = \frac{1}{2}(E_x^\Lambda + E_z^\Lambda - \lambda_\Lambda/2) \pm \frac{1}{2}\sqrt{(E_x^\Lambda - E_z^\Lambda - \lambda_\Lambda/2)^2 + 2\lambda_\Lambda^2}, \tag{1.42b}$$

with each of these being doubly degenerate. We see that $|\text{Bi}_-^+, \pm\frac{1}{2}\rangle$ have the lowest energy out of the Bi$^+$ states, while $|\text{Se}_+^-, \pm\frac{1}{2}\rangle$ have the highest energy among the Se$^-$ states. For strong enough SOC, it can happen that the energy of $|\text{Se}_+^-, \pm\frac{1}{2}\rangle$ is actually higher than that of $|\text{Bi}_-^+, \pm\frac{1}{2}\rangle$, i.e., there exists a level crossing between these states. This corresponds to the the green dashed square of Fig. 1.7 (IV).

The SOC does not only lead to level crossing at the Γ point, but also induces level repulsion between the Bi$_-^+$ and Se$_+^-$ at the touching points. This ultimately leads to an avoided crossing of the valence and conduction bands. Effectively, what has happened with respect to the situation in Fig. 1.7 (III) is that we have interchanged the low-momentum regions of the two bands. The key insight is now that this band swapping takes place between bands of definite but *opposite* parities. At points located far enough from the Γ point, the valence band states are mostly Se$_+^-$ and the conduction band states are mostly Bi$_-^+$, with parity eigenvalues -1 and $+1$, respectively. However, approaching the Γ point mixes these states together, and very close to Γ the valence band is now mostly Bi$_-^+$ and the conduction band is mostly Se$_+^-$. A sketch of the situation can be found in Fig. 1.8. Consequently, moving along one of the bands inverts the parity eigenvalue. Since the Γ point is a time-reversal-invariant momentum, we have changed the sign of one and only one of the parities that enter formula (1.24) for the $\mathbb{Z}_2$ topological invariant[4]. This means that the SOC has converted a trivial insulator into a topological insulator.

---

[4]Note that Eq. (1.24) is only valid for a 2D system, but that a completely analogous expression can be constructed for a three-dimensional topological insulator [30].
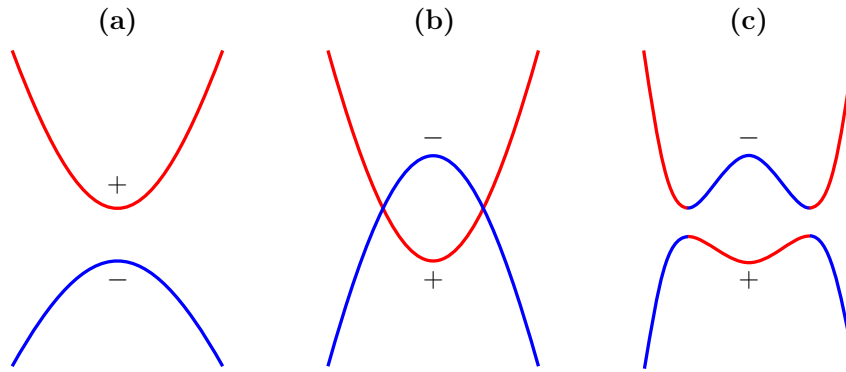
Figure 1.8. Physical effect of the spin-orbit coupling. **(a)** Without SOC, the bands near the Fermi level (assumed to lie inside the gap) maintain the ordering corresponding to completely uncoupled Bi and Se. **(b)** The SOC drives the bands closer to each other, and when the coupling is large enough the energies at the Γ point may invert their ordering. When level repulsion is not taken into account the bands seem to dip into each other; however, this naive picture is unphysical. **(c)** When additionally considering the level repulsion induced by the SOC, the bands feature an avoided crossing with an exchange of parities with respect to the situation in (a).

1.5 A first look at excitons

We have now introduced the first ingredient of this work, namely the concept of a topological insulator and some of its physical manifestations. However, our objective goes beyond the understanding of single-particle states in these materials. Indeed, the goal of this thesis is to study bound electron-hole pairs, known as excitons, in the bulk of Bi_2Se_3 nanosheets. For this reason, in this section we introduce the concept of an exciton and some common terminology. Our purpose here is not to give a detailed mathematical treatment, nor to explain their role in the physics and optical properties of semiconductors. Some excellent and thorough reviews of these vast and intensely researched topics can be found in Refs. [55–59].

1.5.1 The exciton concept

Consider the absorption of a photon by a semiconductor. If the photon energy is large enough, this process can promote a valence electron to the conduction band. A positively charged vacancy, called a *hole*, is left behind in the valence band. The conduction electron and the hole, being oppositely charged, interact electrostatically via the attractive Coulomb interaction. Hence, the two particles can form a bound state which is neutral in charge. This bound electron-hole state is what we call an *exciton*. Being a bound state, the energy of an exciton is lower than the sum of energies of the individual particles. In practice, this means that the energy of the original photon does not necessarily have to be larger than the semiconductor band gap in order to create the exciton.

An exciton carries a crystal pseudomomentum equal to the sum of the individual electron and hole momenta. This pseudomomentum plays the role of a total quasiparticle momentum and allows the exciton to move through the crystal. It also determines the dispersion relation of the exciton when viewed as a single composite particle. In normal semiconductors, where the conduction and valence bands are quadratic with a certain effective mass, the excitonic dispersion is also quadratic in the total momentum. However, in materials with a more complicated band structure, the dispersion relation of excitons may also have a nontrivial dependence on their total momentum. In Ch. 3 we will see that the band inversion of Bi_2Se_3 does indeed have such an effect.

On the other hand, there also exists a relative momentum between the electron and hole that

make up an exciton. This relative momentum determines the internal structure of the exciton. In normal semiconductors, the quadratic dispersion of the single-particle bands allows for the separation of the quantum two-body problem into its center-of-mass and relative components. In this case the internal exciton states are analogous to those of a hydrogen atom with the mass of the proton and electron being substituted by the effective electron and hole masses of the underlying band structure.

In passing, we comment on a potentially confusing point when consulting the literature, where the exciton energy levels are typically represented inside the gap of the semiconductor band diagram. Set the origin of energy at the chemical potential and consider a shift of this quantity, so that the zero of energy is also changed. The individual electron and hole energies are then modified by equal and opposite amounts, which implies that the exciton energy levels are independent of the shift. In other words, the difference between an exciton level and the chemical potential remains constant under any shift of the Fermi level. As a consequence of this, the exciton levels do not actually have a well-defined place in the band diagram. However, when only zero-momentum excitons are studied, it is customary to take the bottom of the conduction band as the position of the electron-hole continuum, i.e., the region without bound electron-hole states, and represent the excitons midgap by subtracting their binding energies. This cannot be done consistently for different values of the total momentum, as the conduction band generally has a different shape from the electron-hole continuum.

1.5.2 Wannier and Frenkel excitons

There are two different types of excitons which can be typically found in semiconductors, and which describe opposite limiting cases. The first type corresponds to tightly bound excitons called *Frenkel excitons*, first introduced in 1931 by Frenkel as "excitation waves" [60]. In Frenkel's model, the interaction between electrons of neighboring atoms is taken to be much weaker than the forces binding each electron to individual atoms. Typically, this is the case when all electronic shells within each atom are closed and the system is insulating, such as in solid xenon or sodium chloride. If one of the electrons in an atom absorbs a quantum of light, it can remain bound to its atom due to the interaction with the positive charge field generated by its own absence from the original closed shell. This describes a very localized exciton whose radius is typically smaller than the size of a single unit cell. Furthermore, the excited atomic state is not confined to a particular atom but can wander through the crystal by passing from one atom to the next. These itinerant excited states are the reason for the name *excitation waves* coined by Frenkel. Due to the reduced size of these kinds of excitons, it is not possible to use the effective electron and hole masses nor the background dielectric constant of the solid for their description.

Clearly, a different approach must be considered to describe excitons in solids where the valence electrons do not longer belong to a single atom, but rather are shared between two or more neighboring ions in the form of covalent bonds. This is typically the situation in semiconductors, where the electrons have gained some room to move in when compared to a fully insulating state. It can be shown that larger interactions between atoms lead to wider valence bands and thus to a larger spatial extent of valence electrons. This scenario is more akin to our above definition of excitons, whereas Frenkel excitons may be regarded as a special case valid for crystals with large band gaps and steep valence bands. The larger freedom of electrons and holes now leads to loosely bound excitons whose spatial extent can range from tens to hundreds of atomic sites. It is now justified to use the background dielectric constant of the solid, ε, to describe the Coulomb interaction between electrons and holes, as well as their effective masses m_e^* and m_h^* arising from the underlying band structure. Hence, one is led to

the *Wannier equation* for a single exciton [61]:

$$\left(-\frac{\boldsymbol{\nabla}_e^2}{2m_e^*} - \frac{\boldsymbol{\nabla}_h^2}{2m_h^*} - \frac{e^2}{4\pi\varepsilon r_{eh}} \right) \Psi_\alpha(\boldsymbol{r}_e, \boldsymbol{r}_h) = E_\alpha \Psi_\alpha(\boldsymbol{r}_e, \boldsymbol{r}_h). \tag{1.43}$$

Here, $\boldsymbol{r}_e$ and $\boldsymbol{r}_h$ are the positions of the electron and the hole, respectively, and $r_{eh} \equiv |\boldsymbol{r}_e - \boldsymbol{r}_h|$. The index α describes the collection of quantum numbers labeling each excitonic state. This model describes so-called *Mott-Wannier* or simply *Wannier excitons*, which are indeed analogous to a hydrogen atom. In particular, the kinetic energy of an exciton with total momentum $\boldsymbol{Q}$ is given by the usual expression

$$K(\boldsymbol{Q}) = \frac{Q^2}{2M_X^*}, \tag{1.44}$$

where $M_X^* = m_e^* + m_h^*$ is the total exciton mass. Note that here we have been dealing with the specialized case of semiconductors with a parabolic band structure. In this work, we will have to consider a more general setting which we will describe in the following chapters. Also, we will only be concerned with Wannier excitons, so we will not consider Frenkel excitons any further.

1.5.3 Bright and dark excitons

In the previous section we have seen how excitons are typically classified according to their size when compared to that of a single unit cell. Another classification exists depending on their response to electromagnetic radiation. *Bright excitons* are those which can be formed simply through the absorption of a single photon. On the other hand, *dark excitons* cannot be created solely via optical transitions and additionally require phonon scattering or some other nonradiative process. Since the momentum carried by a single photon can usually be neglected, bright excitons are typically associated with a direct band gap whereas dark excitons are associated with an indirect gap. This is illustrated in Fig. 1.9.
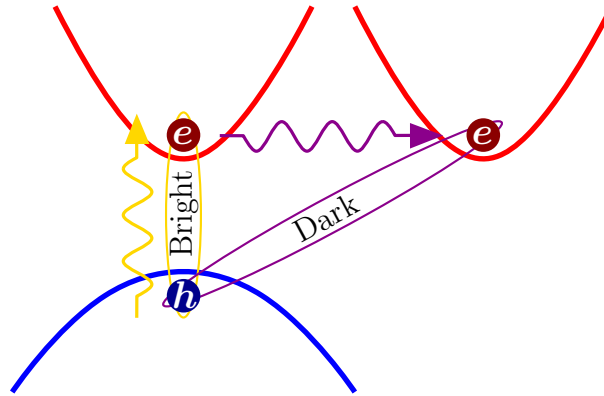

Figure 1.9. Comparison between bright and dark excitons. A bright exciton is typically formed through absorption of a single photon (yellow arrow). By contrast, the electron and the hole in a dark exciton are connected by an optically forbidden transition. The formation of dark excitons thus requires both a photon absorption and a subsequent phonon-scattering event (purple arrow).

The properties of the valence and conduction bands determine the optical generation rate of electron-hole pairs. In quantum mechanics, the dominant contribution to a radiative process is due to electric-dipole radiation. Electric-dipole transitions are subject to the fulfilment of some selection rules. Firstly, the change in total angular momentum J between the initial and final states must satisfy $\Delta J \in \{0, \pm 1\}$, with transitions between two $J = 0$ states being forbidden. Similarly, the total azimuthal angular momentum m_J can only change by $\Delta m_J \in \{0, \pm 1\}$, with

$\Delta m_J \neq 0$ if $\Delta J = 0$ [62, Ch. 9]. Finally, the initial and final states must have opposite parities, since the dipole operator is odd under inversion.

The intensity of optical absorption by excitons in regular semiconductors with quadratic bands was originally studied by Elliott in Ref. [63]. It is found that when dipole transitions are allowed, *s*-wave excitons are bright whereas the rest are dark. Since the spin structure in normal semiconductors leads to a singlet and a triplet, this means that the ground-state singlet is bright whereas the triplet is dark. By contrast, for band structures with dipole-forbidden transitions it is found that *p*-wave excitons are bright and all others are dark. Later on we will study how the situation is modified in Bi_2Se_3. We note that in this material electric-dipole transitions are allowed, because the states $|Bi_-^+, \pm\frac{1}{2}\rangle$ and $|Se_+^-, \pm\frac{1}{2}\rangle$ that make up the valence and conduction bands near the Γ point have opposite parities, and transitions between them can satisfy the angular-momentum rules.

1.5.4 Direct and indirect excitons

Excitons in real materials have a finite lifetime owing to the fact that the electron and the hole can recombine with each other. However, depending on the nature of the excitons, this recombination process may be heavily suppressed. This can lead to excitons with longer lifetimes. For instance, the Wannier model describes excitons with a quadratic kinetic energy whose minimum is located at zero total momentum. This is called a *momentum-direct* exciton. However, we can imagine materials with a more complicated band structure where the kinetic energy of excitons is a more complicated function of the total momentum Q, such that its minimum is shifted away from the origin. These excitons are then called *momentum-indirect* [64, 65]. Momentum-indirect excitons were first observed in silicon and germanium [66]. In such excitons, the individual momenta of the electron and the hole are necessarily different, and thus they cannot easily recombine into a single photon. Therefore, nonradiative processes are required to bring the exciton momentum to zero before radiative recombination is possible. This results in longer-lived excitons.

Directness or indirectness of excitons can also be defined in real space. In a spatially indirect exciton, illustrated in Fig. 1.10, the electron and the hole are confined to different material layers which are spatially separated [67]. Another possibility is to confine the particles in two separate quantum dots [68]. Recombination processes are now supressed because they require interlayer quantum tunneling of one of the composing particles. This can lead to lifetimes several orders of magnitude longer than those of spatially direct excitons.
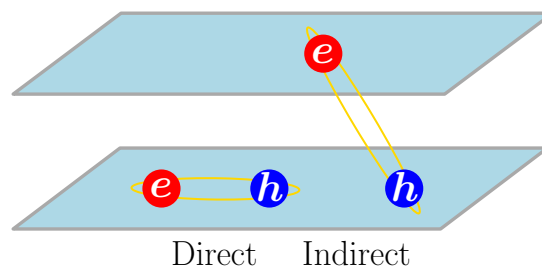

Direct   Indirect

Figure 1.10. Comparison between spatially direct and spatially indirect excitons. In indirect excitons, the electron and the hole are confined to different material layers, and thus recombination is more difficult than for electron-hole pairs living in the same layer.

2

The Low-Energy Model

Having presented the main physics of topological insulators, in this chapter we will lay the foundations of our model for excitons in Bi_2Se_3 nanosheets. We firstly present the specific details of an effective low-energy Hamiltonian suitable for the general description of three-dimensional Bi_2Se_3. We will solve for the eigenstates and energy levels in a thin material slab and give a description of the topological surface states. Then we will integrate out the third dimension and obtain an effective two-dimensional Hamiltonian whose eigenstates and topology will be discussed. This effective two-dimensional Hamiltonian will serve as the basis for our description of excitons in the following chapter.

2.1   Effective Hamiltonian

We are interested in finding an effective Bloch Hamiltonian $H(\mathbf{k})$ describing the states near the Fermi surface in Bi_2Se_3. We have seen that these are $|Bi_-^+, \pm 1/2\rangle$ and $|Se_+^-, \pm 1/2\rangle$. In what follows, we simplify the notation and rename these to $|Bi^+, \sigma\rangle$ and $|Se^-, \sigma\rangle$, respectively, with $\sigma \in \{\uparrow, \downarrow\}$. We choose the ordered basis

$$\left(|Bi^+, \uparrow\rangle, |Se^-, \uparrow\rangle, |Bi^+, \downarrow\rangle, |Se^-, \downarrow\rangle \right). \tag{2.1}$$

In order to describe a particular crystal, the periodic lattice Hamiltonian $\mathcal{H}$ must commute with the operators representing the symmetries of the corresponding lattice. The desired Bloch Hamiltonian is then related to the lattice Hamiltonian through

$$H(\mathbf{k}) = e^{-i\mathbf{k}\cdot\mathbf{r}} \mathcal{H} e^{i\mathbf{k}\cdot\mathbf{r}}, \tag{2.2}$$

and the underlying symmetries are expressed through specific relations between the Bloch matrices at different momenta. For instance, the Bi_2Se_3 lattice symmetries are realized as

$$\mathbf{R}H(\mathbf{k})\mathbf{R}^{-1} = H(R_3\mathbf{k}), \tag{2.3a}$$

$$\mathbf{P}H(\mathbf{k})\mathbf{P}^{-1} = H(-\mathbf{k}), \tag{2.3b}$$

$$\mathbf{T}H(\mathbf{k})\mathbf{T}^{-1} = H(-\mathbf{k}), \tag{2.3c}$$

where R, P, and T are appropriate representations of the three-fold rotation, parity, and time-reversal symmetries, respectively, acting on the combined spin and orbital space (2.1). Also, $R_3 \equiv R_z(2\pi/3)$, with $R_z(\theta)$ the usual rotation matrix around the z-axis. Note that, even though P and T act on $H(k)$ in the same way, the former operator is unitary whereas the latter is antiunitary. In our basis, the representations of these symmetry operations are [53]

$$R = \exp[i(\pi/3)\sigma_z \otimes \mathbb{I}_2], \tag{2.4a}$$
$$P = \mathbb{I}_2 \otimes \tau_z, \tag{2.4b}$$
$$T = i\sigma_y \otimes \mathbb{I}_2 \, \mathcal{K}, \tag{2.4c}$$

where σ_i and τ_i are the Pauli matrices acting on the spin ($|\uparrow\rangle$, $|\downarrow\rangle$) and orbital ($|\mathrm{Bi}^+\rangle$, $|\mathrm{Se}^-\rangle$) spaces, respectively. Again, $\mathcal{K}$ denotes the complex conjugation operator, with $\mathcal{K}i\mathcal{K}^{-1} = -i$ and $\mathcal{K}^2 = 1$. It is immediate to verify that $T^2 = -1$.

An effective Bloch Hamiltonian possessing a given set of crystal symmetries can always be derived in the vicinity of an arbitrary crystal momentum k_0 by means of the so-called $k \cdot p$ *method* [69, Ch. 2, 5, Ch. 9]. The most general three-dimensional Hamiltonian compatible with the symmetries of $\mathrm{Bi_2Se_3}$ is derived from $k \cdot p$ theory in Refs. [53, 54] around the Γ point. To quadratic order in k, it reads

$$H(k) = \epsilon_0(k) + \mathcal{M}(k)\,\mathbb{I}_2 \otimes \tau_z + A_1 k_z\, \sigma_z \otimes \tau_x + A_2(k_x\sigma_x + k_y\sigma_y) \otimes \tau_x, \tag{2.5}$$

where

$$\epsilon_0(k) = C + D_1 k_z^2 + D_2(k_x^2 + k_y^2), \tag{2.6a}$$
$$\mathcal{M}(k) = M - B_1 k_z^2 - B_2(k_x^2 + k_y^2). \tag{2.6b}$$

In matrix form, it is explicitly given by

$$H(k) = \epsilon_0(k)\,\mathbb{I}_4 + \begin{bmatrix} \mathcal{M}(k) & A_1 k_z & 0 & A_2 k_- \\ A_1 k_z & -\mathcal{M}(k) & A_2 k_- & 0 \\ 0 & A_2 k_+ & \mathcal{M}(k) & -A_1 k_z \\ A_2 k_+ & 0 & -A_1 k_z & -\mathcal{M}(k) \end{bmatrix}, \tag{2.7}$$

with $k_\pm = k_x \pm i k_y$. The parameter values for $\mathrm{Bi_2Se_3}$ are obtained by fitting existing experimental data to the theoretical model and are given in Table 2.1.

Parameter	Value
C	$-0.0068\,\mathrm{eV}$
M	$0.28\,\mathrm{eV}$
A_1	$0.22\,\mathrm{eV\,nm}$
A_2	$0.41\,\mathrm{eV\,nm}$
B_1	$0.10\,\mathrm{eV\,nm^2}$
B_2	$0.566\,\mathrm{eV\,nm^2}$
D_1	$0.013\,\mathrm{eV\,nm^2}$
D_2	$0.196\,\mathrm{eV\,nm^2}$

Table 2.1. Model parameters for the $k \cdot p$ Hamiltonian for three-dimensional $\mathrm{Bi_2Se_3}$ of Refs. [53, 54]. These values are extracted in Ref. [53] by fitting the resulting band structure to experimental data.

The above Hamiltonian belongs to class AII in the tenfold way of Table 1.2. The three-dimensional $\mathbb{Z}_2$ topological invariant, which is also obtained via a product of parities as in

the 2D case, indicates that it is a topological insulator. Note that it is in fact identical (modulo a unitary transformation) to the Hamiltonian for a 3D topological insulator presented in Ref. [70]. The $\mathbb{Z}_2$ topological invariant is explicitly evaluated in Ref. [53] for several materials, and is nontrivial in Bi_2Se_3. This implies that we expect boundary modes at the surfaces of a thin two-dimensional sample. These surface states will be derived explicitly in the next section.

It is important to note that, in our effective theory, we let momenta become continuous. This implies that we effectively send the lattice spacing to zero and thus get rid of the Brillouin zone and the periodicity of the Bloch Hamiltonian, which we from now on refer to as simply the Hamiltonian. However, the Hamiltonian still describes the Bi_2Se_3 crystal in the sense that it still exhibits the symmetries of the original lattice.

In this thesis we shall be concerned with effectively two-dimensional material samples. For this reason, from now on we redefine $x \equiv (x, y)$ and $k \equiv (k_x, k_y)$ as the in-plane coordinates and momenta, respectively, and additionally indicate the z-dependence exlicitly when necessary. For instance, our Hamiltonian is now written as $H(k, k_z)$, and the Γ point is now simply defined as $k_x = k_y = 0$.

2.2 Solution of the 3D model

Since we are interested in the description of Bi_2Se_3 nanosheets, in this section we will solve the Hamiltonian (2.7) in a thin material slab of thickness ℓ lying on the xy-plane. We closely follow the method of Refs. [71, 72]. The in-plane momenta k_x and k_y are still good quantum numbers, and we choose to solve the model at the Γ point because we are merely interested in the low-energy physics. The solution for arbitrary k can be obtained through a similar procedure and is derived in Ref. [73]. In our case, the Hamiltonian reduces to a block-diagonal form,

$$H(0, k_z) = \begin{bmatrix} h_+(k_z) & 0 \\ 0 & h_-(k_z) \end{bmatrix},$$

(2.8)

where $h_\pm(k_z)$ are 2×2 matrices given by

$$h_\pm(k_z) = (C + D_1 k_z^2) + (M - B_1 k_z^2)\sigma_z \pm A_1 k_z \sigma_x.$$

(2.9)

In view of our choice of basis (2.1), h_+ and h_- describe fermions with a well-defined spin in the z-direction ($\uparrow$ and $\downarrow$, respectively). On the other hand, k_z is not a good quantum number anymore due to the breaking of translational invariance along the z-direction caused by the geometry of the slab. For this reason we perform the substitution $k_z \to -i\partial_z$. Now we can solve either of these 2×2 Hamiltonians, say h_+, by proposing a plane-wave ansatz for the wave function, $\psi_\Lambda(z) \propto u_\Lambda e^{\Lambda z}$, such that $h_+(-i\partial_z)\psi_\Lambda(z) = E\psi_\Lambda(z)$ for some trial energy E. Here, u_Λ is a (z-independent) two-component vector satisfying the eigenvalue equation $h_+(-i\Lambda)u_\Lambda = Eu_\Lambda$. Naturally, there may be several solutions for Λ, in which case the total wave function $\Psi(z)$ will be a linear superposition of the different $\psi_\Lambda(z)$ with coefficients fixed by boundary conditions. In our case, we will assume that the electronic states do not leak out of the nanosheet, and thus we choose the hard-wall boundary conditions $\Psi(z = \pm\ell/2) = 0$.

The matrix $h_+(-i\Lambda)$ has two eigenvalues, namely

$$\lambda_\pm = C - D_1\Lambda^2 \pm \sqrt{(B_1\Lambda^2 - M)^2 - A_1\Lambda^2}.$$

(2.10)

Setting either of them equal to E gives four distinct solutions for Λ, which come in the form $\pm\Lambda_\pm(E)$, where

$$\Lambda_\pm(E) = \sqrt{\frac{-R(E) \pm \sqrt{S(E)}}{2(D_1^2 - B_1^2)}}.$$

(2.11)

Here we have defined

$$R(E) \equiv A_1^2 + 2D_1(E - C) - 2B_1 M, \tag{2.12a}$$

$$S(E) \equiv R(E)^2 - 4(D_1^2 - B_1^2)[(E - C)^2 - M^2]. \tag{2.12b}$$

In principle one would have to be careful not to introduce spurious solutions for Λ, which is a possibility due to the nonlinear nature of Eq. (2.10). In this particular case, however, it turns out that the spurious solutions of the equation $\lambda_+ = E$ are actually the true solutions of $\lambda_- = E$, and vice versa. Since we are interested in satisfying either of the equations, as they both impose that E be an eigenvalue of the Hamiltonian, we can ultimately keep all the solutions in Eq. (2.11). Now let Λ be a solution to either $\lambda_+ = E$ or $\lambda_- = E$. In both cases, the corresponding eigenvector of the matrix $h_+(-i\Lambda)$ reads

$$\boldsymbol{u}_\Lambda = \begin{bmatrix} g(\Lambda) \\ -iA_1\Lambda \end{bmatrix}, \tag{2.13}$$

where $g(\Lambda) \equiv E + M - C + (B_1 + D_1)\Lambda^2$. Note that we assume that A_1 is nonzero, otherwise $h_+(k_z)$ is already diagonal and equivalent to two uncoupled infinite wells. Since the eigenstates read $\psi_\Lambda(z) = \boldsymbol{u}_\Lambda e^{\Lambda z}$, and both Λ and $-\Lambda$ are allowed, we may now create wave functions with definite-parity components, $\psi_\Lambda^\pm(z) \propto \psi_\Lambda(z) \pm \psi_{-\Lambda}(z)$, and from now on $\Lambda \in \{\Lambda_+, \Lambda_-\}$ only. Then we obtain the solutions

$$\psi_\Lambda^+(z) = \begin{bmatrix} g(\Lambda)\cosh(\Lambda z) \\ -iA_1\Lambda\sinh(\Lambda z) \end{bmatrix}, \qquad \psi_\Lambda^-(z) = \begin{bmatrix} g(\Lambda)\sinh(\Lambda z) \\ -iA_1\Lambda\cosh(\Lambda z) \end{bmatrix}. \tag{2.14}$$

Finally, we now see that there are two eigenstates, namely $\varphi_+(z) \equiv \sum_{\Lambda \in \{\Lambda_+,\Lambda_-\}} \alpha_\Lambda \psi_\Lambda^-(z)$ and $\chi_+(z) \equiv \sum_{\Lambda \in \{\Lambda_+,\Lambda_-\}} \beta_\Lambda \psi_\Lambda^+(z)$, as the boundary conditions at $\pm\ell/2$ are equivalent due to the well-defined parity of all components. The subscript $+$ in φ_+ and χ_+ stands for the fact that these are eigenstates of h_+. The eigenstates of h_- are similarly denoted by φ_- and χ_-, and are simply obtained by changing A_1 to $-A_1$ in φ_+ and χ_+, respectively. Requiring a nontrivial solution for the coefficients α_Λ and β_Λ leads to two transcendental equations, namely

$$\frac{[E + M - C + (B_1 + D_1)\Lambda_+^2]\Lambda_-}{[E + M - C + (B_1 + D_1)\Lambda_-^2]\Lambda_+} = \frac{\tanh(\Lambda_-\ell/2)}{\tanh(\Lambda_+\ell/2)} \quad \text{for } \varphi_+, \tag{2.15a}$$

$$\frac{[E + M - C + (B_1 + D_1)\Lambda_+^2]\Lambda_-}{[E + M - C + (B_1 + D_1)\Lambda_-^2]\Lambda_+} = \frac{\tanh(\Lambda_+\ell/2)}{\tanh(\Lambda_-\ell/2)} \quad \text{for } \chi_+. \tag{2.15b}$$

These can be solved numerically for the allowed energies of the states φ_+ and χ_+, which we denote by E_φ and E_χ, respectively. There are in general many solutions, corresponding to the quantized energy levels of the problem under consideration. The eigenstates are

$$\varphi_+(z) = \mathcal{N}_\varphi \begin{bmatrix} (B_1 + D_1)\eta_\varphi S_\varphi(z) \\ -iA_1 C_\varphi(z) \end{bmatrix} \quad \text{with energy } E_\varphi, \tag{2.16a}$$

$$\chi_+(z) = \mathcal{N}_\chi \begin{bmatrix} (B_1 + D_1)\eta_\chi C_\chi(z) \\ -iA_1 S_\chi(z) \end{bmatrix} \quad \text{with energy } E_\chi, \tag{2.16b}$$

with $\mathcal{N}_\varphi$ and $\mathcal{N}_\chi$ normalization constants and

$$C_\xi(z) = \frac{\cosh(\Lambda_+^\xi z)}{\cosh(\Lambda_+^\xi \ell/2)} - \frac{\cosh(\Lambda_-^\xi z)}{\cosh(\Lambda_-^\xi \ell/2)}, \tag{2.17a}$$

$$S_\xi(z) = \frac{\sinh(\Lambda_+^\xi z)}{\sinh(\Lambda_+^\xi \ell/2)} - \frac{\sinh(\Lambda_-^\xi z)}{\sinh(\Lambda_-^\xi \ell/2)}, \tag{2.17b}$$

$$\eta_\varphi = \frac{(\Lambda_+^\varphi)^2 - (\Lambda_-^\varphi)^2}{\Lambda_+^\varphi \coth(\Lambda_+^\varphi \ell/2) - \Lambda_-^\varphi \coth(\Lambda_-^\varphi \ell/2)}, \tag{2.17c}$$

$$\eta_\chi = \frac{(\Lambda_+^\chi)^2 - (\Lambda_-^\chi)^2}{\Lambda_+^\chi \tanh(\Lambda_+^\chi \ell/2) - \Lambda_-^\chi \tanh(\Lambda_-^\chi \ell/2)}. \tag{2.17d}$$

Here we have defined $\Lambda_\pm^\xi \equiv \Lambda_\pm(E_\xi)$, with $\xi \in \{\varphi, \chi\}$, which can be found by putting the solution E_ξ obtained from Eqs. (2.15) back into (2.11). As we have said, the eigenstates of h_-, φ_- and χ_-, are obtained from Eqs. (2.16) by changing A_1 to $-A_1$. In particular, they are degenerate in energy with φ_+ and χ_+, respectively. This is an immediate consequence of the Kramers degeneracy theorem.

With this, we have a complete description of the eigenstates of the $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian at the Γ point. In the following sections we will show examples of the resulting probability densities, and in particular we will show that the spectrum includes localized surface states due to the nontrivial topology.

2.2.1 Bulk states

By solving the transcendental equations (2.15) we firstly obtain energy levels corresponding to the usual semiconductor bulk bands at the Γ point. Of course, the results will depend on the thickness of the nanosheet, which is our only free parameter. For reasons that will become clear below, we choose $\ell = 6$ nm in this section. The eigenenergies for the first few states are shown in Table 2.2. Also, Fig. 2.1 shows the probability densities of some of these states as computed from Eqs. (2.16).

n	E_c (eV)	E_v (eV)
1	0.281	−0.284
2	0.342	−0.308
3	0.503	−0.412
4	0.756	−0.596
5	1.084	−0.843
6	1.480	−1.146

Table 2.2. Energies of several low-lying conduction and valence states in the bulk of a thin Bi_2Se_3 nanosheet of thickness $\ell = 6$ nm, as obtained from Eqs. (2.15). All energies are measured with respect to the Fermi level.

One may object that the value of $E_c - E_v = 0.565$ eV (for $n = 1$) overestimates the band gap of Bi_2Se_3, as theoretical studies report values of 0.31 eV–0.32 eV [74, 75] and experimental data gives 0.35 eV [76]. However, $E_c - E_v$ is not necessarily expected to match these values. The reason is that, at least in our model, Bi_2Se_3 is an indirect semiconductor; in other words, the band gap is *not* the difference between the energies at the Γ point. In fact, the model of Ref. [53] on which the Hamiltonian (2.7) is based *does* predict an indirect band gap of 0.3 eV, as well as our obtained Γ-point gap of ~ 0.565 eV. Furthermore, although to much lesser degree, the literature values for the band gap presented here correspond to *bulk* Bi_2Se_3, whereas the thin slab geometry could in principle alter the bands. In practice, however, this effect only seems relevant for a number of QLs smaller than that corresponding to our choice of ℓ [75, 77].
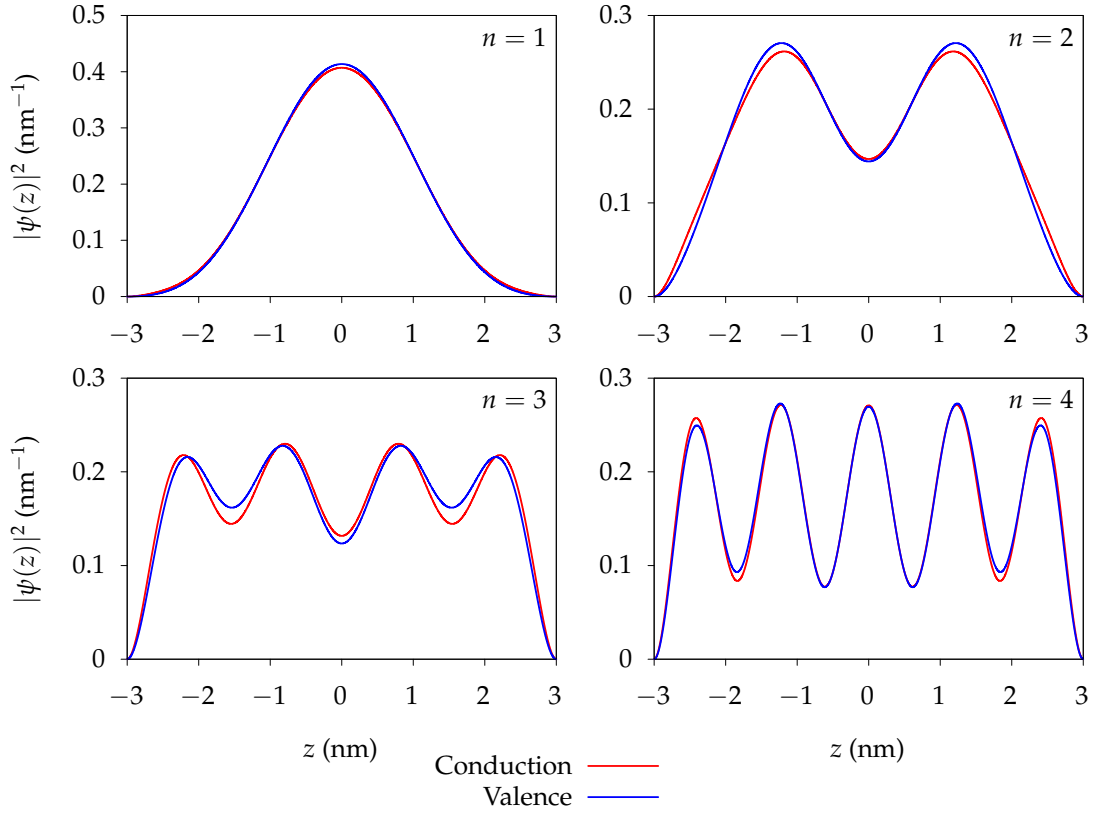
Figure 2.1. Probability densities of the lowest four conduction (red) and valence (blue) states in the bulk of a thin Bi_2Se_3 nanosheet of thickness $\ell = 6\,\text{nm}$, as obtained from Eqs. (2.16). The corresponding energies are found in Table 2.2. We note that, due to the vector nature of the eigenstates, the probability densities do not exhibit the characteristic nodes of the infinite-well solutions.

2.2.2 Surface states

Arguably, one of the most interesting features of the spectrum we have obtained is the presence of topologically protected surface states. When the thickness ℓ is large enough, one finds that the transcendental equations (2.15) contain a common solution, $E_\varphi = E_\chi$, which lies in the bulk band gap of the insulator. This is the so-called *Dirac point*. We denote the surface-state eigenfunctions of $h_\pm$ by $\varphi_\pm^S$ and $\chi_\pm^S$. Plotting the probability density of any of the four degenerate states in this case reveals that these are localized near the two surfaces of the slab. As we reduce ℓ, quantum tunneling processes between the top and bottom surfaces become possible and a small gap $\Delta \equiv |E_\varphi - E_\chi|$ opens at the Γ point [71–73, 78, 79]. The probability densities, which up to this point were identical for all surface states, now become slightly different. This effect will be discussed in more detail below. In Fig. 2.2, where we have plotted the probability densities for several values of ℓ, all these features can be observed. Finally, note that the fact that $E_\varphi \neq E_\chi$ does not contradict the Kramers theorem, as there are still two states for each value of the energy.

For later convenience, we now define the following ordered basis spanning the surface-state subspace for a given thickness:

$$(\psi_1, \psi_2, \psi_3, \psi_4) \equiv \left(\begin{bmatrix} \varphi_+^S(z) \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \chi_-^S(z) \end{bmatrix}, \begin{bmatrix} \chi_+^S(z) \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \varphi_-^S(z) \end{bmatrix} \right). \tag{2.18}$$

The probability densities of these eigenstates are symmetric about $z = 0$. However, it will often be convenient to create linear combinations of these states such that the probability density is

30

localized near one of the two surfaces. One may verify that, in the absence of a gap, not only $\Lambda_\pm^\varphi = \Lambda_\pm^\chi$, but also $\eta_\varphi = \eta_\chi$ and $\mathcal{N}_\varphi = \mathcal{N}_\chi$ in Eqs. (2.16). Then, by defining

$$\psi_{t,\uparrow} \equiv \frac{1}{\sqrt{2}}(\psi_1 + \psi_3), \qquad \psi_{t,\downarrow} \equiv \frac{1}{\sqrt{2}}(\psi_2 + \psi_4), \tag{2.19a}$$

$$\psi_{b,\uparrow} \equiv \frac{1}{\sqrt{2}}(\psi_1 - \psi_3), \qquad \psi_{b,\downarrow} \equiv \frac{1}{\sqrt{2}}(\psi_2 - \psi_4), \tag{2.19b}$$

we obtain eigenstates of the Hamiltonian whose probability densities are localized near the top ($z = \ell/2$) or bottom ($z = -\ell/2$) surfaces, as indicated by the subscripts t and b, respectively. Furthermore, the spin in the z-direction remains well-defined because we are not mixing the eigenstates of h_+ and h_-. Another advantage which will be important later is that these states are orthogonal even *without* the integral over z, i.e.,

$$\psi_{s,\sigma}^\dagger(z)\psi_{s',\sigma'}(z) \propto \delta_{s,s'}\delta_{\sigma,\sigma'}. \tag{2.20}$$

We stress that, strictly speaking, these results are only valid when $\Delta = 0$. However, we have numerically verified that they still hold to good approximation when Δ is relatively small, in practice all the way down to $\ell \approx 6\,\mathrm{nm}$. Reasons for this are given in the following section, where we discuss the behavior in the ultrathin limit.
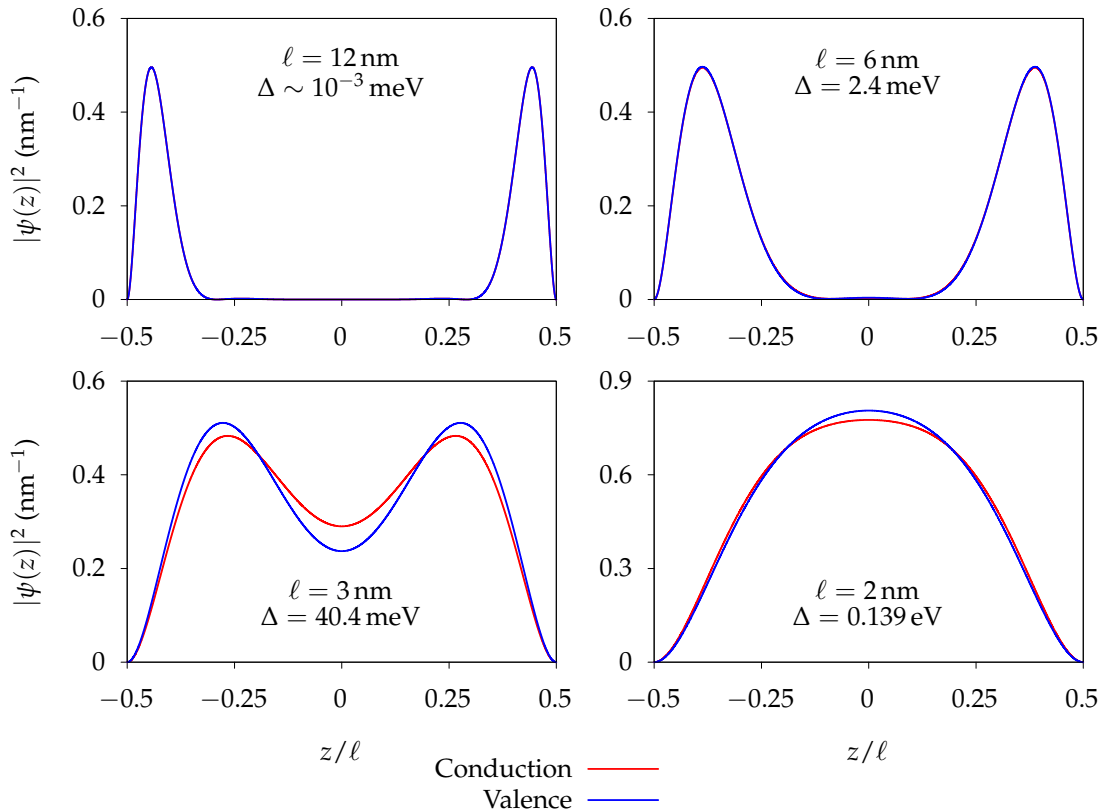

Figure 2.2. Probability densities of the conduction (red) and valence (blue) surface states of a thin Bi$_2$Se$_3$ nanosheet of varying thickness, as obtained from Eqs. (2.16). For large ℓ, the states are localized at the surfaces and have the same energy. As the thickness decreases, tunneling processes between the two surfaces become possible and the probability densities become more and more bulk-like. The coupling between the surfaces opens a gap Δ which breaks the degeneracy of these states at the Γ point. The Kramers theorem is preserved because in all cases there are actually two states of each type.

2.2.3 Effects of material thickness

As we have said, when ℓ is large enough the system hosts gapless, topologically robust surface states. Decreasing the thickness eventually produces a hybridization between the modes at both surfaces. This overlap induces level repulsion and a nonzero energy gap opens at the Dirac point. The analytical solution of the model presented in the previous section does not show an observable gap when the thickness is larger than around 6 nm. This agrees with experimental data [80, 81] and density functional theory calculations [82] showing that Bi_2Se_3 remains ungapped all the way down to six quintuple layers, as the thickness of one QL is around 1 nm [83–85].

A more careful analysis shows that the decay of the gap with increasing ℓ is nonmonotonous, and features an exponential decay with a superimposed oscillatory pattern [71, 73, 79]. In this picture, the material alternates between a trivial and a topological quantum spin Hall phase. This oscillatory behavior, however, has yet to be observed experimentally. In fact, the results of Ref. [80] suggest that the period of these oscillations might be too large to observe them before the system becomes essentially gapless for growing ℓ. Moreover, there seems to be some disagreement in the literature regarding the nature of the phase below the 6 QL threshold. While the theoretical predictions of [71] in combination with the experimental findings of [80] predict that a quantum spin Hall phase should exist between 2 and 5 QLs, Ref. [81] claims that in this regime Bi_2Se_3 behaves like a trivial insulator.

In any case, all theoretical and experimental studies agree that for 6 QLs or more, which corresponds to a thickness of $\ell \gtrsim 6$ nm, the material is in a topologically nontrivial phase. Furthermore, the asymptotic large-ℓ behavior is recovered extremely quickly with increasing number of QLs, which explains why Eqs. (2.19) and (2.20) are still valid to a very good approximation even when approaching the critical thickness. Topological protection of the surface states is guaranteed in the large-thickness regime and has also been observed experimentally in Ref. [86], where it is demonstrated that for $\ell \gtrsim 6$ nm the system exhibits metallic surface transport. By contrast, the surface conductance abruptly drops below this critical thickness.

2.3 Effective surface Hamiltonian

So far we have only solved the model at the two-dimensional Γ point, which prevents us from seeing the shape of the dispersions for nonzero values of the momentum. However, knowledge of the surface states $\psi_\alpha(z)$ at the Γ point is enough to derive an effective low-energy Hamiltonian. This Hamiltonian will be denoted by $H_S(\boldsymbol{k})$ and will allow us to obtain the correct dispersion relation for small but nonzero values of $\boldsymbol{k}$. The components of H_S are obtained by projecting $H(\boldsymbol{k}, -\mathrm{i}\partial_z)$ onto the Hilbert space spanned by the four surface states [54, 71, 72]:

$$H_S(\boldsymbol{k})_{\alpha\beta} = \langle\psi_\alpha|\, H(\boldsymbol{k}, -\mathrm{i}\partial_z)\, |\psi_\beta\rangle = \int_{-\ell/2}^{\ell/2} \mathrm{d}z\, \psi_\alpha^\dagger(z) H(\boldsymbol{k}, -\mathrm{i}\partial_z)\psi_\beta(z)\,. \tag{2.21}$$

In the basis $(\psi_1, \psi_2, \psi_3, \psi_4)$ this always turns out to be a 4×4 block-diagonal matrix with two 2×2 entries. Furthermore, when $\Delta = 0$ this is also the case in the basis $(\psi_{t,\uparrow}, \psi_{t,\downarrow}, \psi_{b,\uparrow}, \psi_{b,\downarrow})$, and one may verify that the components in both bases are equal. For this reason, in the absence of a gap the two 2×2 blocks can be seen as describing the top and bottom surfaces separately. As expected, when the Dirac point becomes gapped, the Hamiltonian in the latter basis additionally contains a small coupling between the equal-spin states at the top and bottom surfaces. We will henceforth assume the gapless limit at all times. This is a good approximation in the continuous model even for $\ell = 6$ nm, and experimentally holds down to 6 quintuple

layers. In this case, both 2×2 blocks are identical and read

$$h_{\mathrm{S}}(\boldsymbol{k}) = E_0 + D|\boldsymbol{k}|^2 - v_{\mathrm{F}}(\boldsymbol{k} \times \boldsymbol{\sigma}) \cdot \hat{\boldsymbol{z}}. \tag{2.22}$$

Here E_0 is the energy at the Γ point, and we take the asymptotic (i.e., large-ℓ) values for the various parameters. Their values are shown in Table 2.3. The last term, which explicitly reads $(\boldsymbol{k} \times \boldsymbol{\sigma}) \cdot \hat{\boldsymbol{z}} = k_x \sigma_y - k_y \sigma_x$, is called a *Rashba term*. It arises from the combined effect of the spin-orbit coupling and the asymmetry of the crystal potential in the directions parallel and perpendicular to the nanosheet [87]. A more detailed analysis, such as that of Ref. [54], would reveal higher-order terms proportional to k^3 that break rotational symmetry. For small enough ℓ, the expected gap term is also reproduced [71], even though we have neglected it in our large-thickness limit.

Parameter	Value
E_0	$0.0296\,\mathrm{eV}$
D	$0.122\,\mathrm{eV\,nm^2}$
v_{F}	$0.407\,\mathrm{eV\,nm}$

Table 2.3. Asymptotic values for the parameters of the effective Bi_2Se_3 surface Hamiltonian.

The eigenenergies of this Hamiltonian are

$$\epsilon_{\pm}(\boldsymbol{k}) = E_0 \pm v_{\mathrm{F}}|\boldsymbol{k}| + D|\boldsymbol{k}|^2. \tag{2.23}$$

For small momenta the energies grow linearly with $|\boldsymbol{k}|$, as corresponds to a model of massless particles, in this case Dirac fermions. The parameter v_{F} is called the *Fermi velocity*. At zero momentum the valence and conduction bands cross linearly at a single point and form the characteristic *Dirac cone*. The dispersions of the surface states for small $\boldsymbol{k}$ are shown in Fig. 2.3, where we also show the first valence and conduction bulk states determined in the following section.

Due to the Rashba term in h_{S}, the surface eigenstates always satisfy $\langle \sigma_z \rangle = 0$. This does not contradict our discussion of the two-dimensional QSHE, as the surface modes derived here correspond to a *three*-dimensional topological insulator. Quite on the contrary, the physical picture is the same, as the phenomenon of spin-momentum locking remains. Indeed, the expected values of the spin operator in the states of energy $\epsilon_{\pm}(\boldsymbol{k})$ are

$$\langle \boldsymbol{\sigma} \rangle_{\pm} = \pm \cos \phi_k \, \hat{\boldsymbol{x}} \pm \sin \phi_k \, \hat{\boldsymbol{y}}, \tag{2.24}$$

where ϕ_k is the polar angle of the momentum $\boldsymbol{k}$ with the x-axis. This equation implies that the electron's spin always tends to form a right angle with the direction of its momentum, winding exactly once around itself as ϕ_k changes from 0 to 2π. Spin-momentum locking is the source of robustness of these surface states. If the system contains impurities that do not break the time-reversal symmetry, then these impurities cannot change the direction of the spin. And since the direction of $\boldsymbol{k}$ is locked to that of the spin, it follows that scattering events due to these impurities are strongly suppressed, i.e., they will not be able to change the electron's direction of motion.

2.4 Effective bulk Hamiltonian

The procedure we have followed in order to obtain an effective surface Hamiltonian can equally well be used for the bulk. This effective bulk Hamiltonian will be our starting point for the treatment of excitons in the next chapter. However, there is one difference with the surface

case, which is that the number of bulk states one obtains in our continuous model is infinite, since Eqs. (2.15) give infinitely many solutions. The complete effective bulk Hamiltonian would then of course also be a matrix with infinite entries. However, let us for the moment imagine that the low-energy physics of the problem of interest are somehow constrained to the first valence and conduction bands, i.e., transitions to higher excited states are not allowed. Then it is correct to consider only the corresponding subspace of the effective Hamiltonian. We will assume that this is the case for our exciton problem of the following chapter. For this reason, we proceed to project the three-dimensional Hamiltonian onto the subspace of the single-particle Hilbert space spanned *only* by the first conduction and valence subbands, and neglect the effect of all the higher excited states. In the next section we will comment on the validity of this approximation, which will be ultimately verified in Ch. 3.

Solving the model at the Γ point for the bulk conduction and valence ground states as explained in Sec. 2.2, we find two lowest-lying conduction ground states $\chi_{\pm}^c$ and two highest-lying valence ground states $\varphi_{\pm}^v$. We then choose the following ordered basis of eigenfunctions:

$$\left(\begin{bmatrix} \chi_+^c(z) \\ 0 \end{bmatrix}, \begin{bmatrix} i\varphi_+^v(z) \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ \chi_-^c(z) \end{bmatrix}, \begin{bmatrix} 0 \\ -i\varphi_-^v(z) \end{bmatrix} \right). \tag{2.25}$$

The reason for the factors of $\pm i$ will become clear in a moment. Projecting the Hamiltonian onto the subspace spanned by these eigenfunctions in the same way of Eq. (2.21) gives an effective, z-independent Hamiltonian for the bulk. It reads

$$\widetilde{H}(\boldsymbol{k}) = \widetilde{\epsilon}_0(\boldsymbol{k})\,\mathbb{I}_4 + \begin{bmatrix} \widetilde{\mathcal{M}}(\boldsymbol{k}) & 0 & 0 & \widetilde{A}_2 k_- \\ 0 & -\widetilde{\mathcal{M}}(\boldsymbol{k}) & \widetilde{A}_2 k_- & 0 \\ 0 & \widetilde{A}_2 k_+ & \widetilde{\mathcal{M}}(\boldsymbol{k}) & 0 \\ \widetilde{A}_2 k_+ & 0 & 0 & -\widetilde{\mathcal{M}}(\boldsymbol{k}) \end{bmatrix}, \tag{2.26}$$

where $\widetilde{\epsilon}_0(\boldsymbol{k}) = \widetilde{C} + \widetilde{D}_2|\boldsymbol{k}|^2$ and $\widetilde{\mathcal{M}}(\boldsymbol{k}) = \widetilde{M} - \widetilde{B}_2|\boldsymbol{k}|^2$. We see that this Hamiltonian is precisely of the form of (2.7), with $A_1 = B_1 = D_1 = 0$ and with some renormalized values of the model parameters. We have labeled these with a tilde in the equation above, but we shall drop it henceforth. The values of the renormalized parameters are shown in Table 2.4. The global phases in Eq. (2.25) can now be easily understood; we have chosen them so that our effective Hamiltonian has exactly the same form as its three-dimensional counterpart instead of being related to it by a unitary transformation. One small caveat to this is that the basis in which the effective Hamiltonian is expressed is *not* that of Eq. (2.1), because the Bi^+ and Se^+ orbitals are hybridized in the eigenstates (2.16). Of course, this does not influence the emerging physics in any way.

Parameter	Bare (3D model)	Renormalized ($\ell = 6$ nm)
C	-0.0068 eV	-0.0012 eV
M	0.28 eV	0.28 eV
A_2	0.41 eV nm	0.41 eV nm
B_2	0.566 eV nm^2	0.473 eV nm^2
D_2	0.196 eV nm^2	0.202 eV nm^2

Table 2.4. Bare parameters for the three-dimensional model and their renormalized values after projecting the Hamiltonian onto the single-particle ground states for $\ell = 6$ nm.

We have now integrated out the z-dependence and obtained an effective model suitable for describing the physics inside a thin material nanosheet. The above Hamiltonian will be valid for problems which can be mostly regarded as two-dimensional in the sense that the

dynamics in the *z*-direction are largely unimportant in comparison to the in-plane physics. In the following we will work with the renormalized values of the material parameters in order to describe the bulk of our Bi_2Se_3 nanosheets. Our numerical values for exciton binding energies, radii, etc., which we will obtain in Ch. 3, will thus be valid for the particular value $\ell = 6\,\mathrm{nm}$. However, all *topological* arguments and conclusions will still be valid for any value of ℓ small enough for the problem to be regarded as two-dimensional, since a continuous deformation of ℓ would merely produce a change in the renormalization of the model parameters.

2.4.1 A word on the first-band approximation

Before moving on, let us comment on our *first-band approximation* consisting in disregarding the excited bulk states. In general we expect this approximation to become more accurate with decreasing film thickness, as the energy splitting of the subbands in a naive infinite-well model goes as $m^{-1}\ell^{-2}$, with m the effective mass of the particle. Even so, in the particular case of Bi_2Se_3 the effective masses are relatively small, as can be checked by diagonalizing (2.7) as a function of k_z at the 2D Γ point; this yields bands that are rather flat at the origin. For this reason, it is not entirely clear *a priori* that the physics of an interacting system should play out almost exclusively in this subspace. However, let us particularize to our problem of interest, namely that of finding bound electron-hole eigenstates of a certain interacting two-body Hamiltonian. The interaction potential between the two particles will induce transitions between energy levels whose splitting is smaller than their binding energy, which is the characteristic energy available for the exchange of a virtual photon. Hence, we expect the effect of bands whose energy differs from that of the first band by an energy larger than the exciton binding energy to be negligible. This gives a quantitative way of checking the correctness of our approximation *a posteriori*: it will be valid whenever the obtained binding energies are smaller than the splitting of the $n = 1$ and $n = 2$ states of Table 2.2.

On another note, we must mention that we have also quietly neglected transitions between the first conduction and valence bands and the surface states. Of course, a global tight-binding Hamiltonian would need to include a hopping term between these subspaces. However, we have neglected this term in view of the fact that it depends on the overlap of wave functions. Since the surface states are localized at both sides of the sample, while the bulk wave functions are maximized around the middle, this overlap is expected to be small. This is precisely the statement that the surface states are robust: their strong coherence makes hopping to other states unlikely.

2.5 Analysis of the bulk Hamiltonian

In this section we study the physics behind our newly obtained effective bulk Hamiltonian. We will first give expressions for the energy bands and eigenstates, and then we will give a careful explanation of the nontrivial topological properties that it gives rise to.

2.5.1 Band structure and eigenstates

The Hamiltonian (2.26) can be straightforwardly diagonalized and gives the eigenenergies

$$\epsilon_{c,v}(\boldsymbol{k}) = \epsilon_0(\boldsymbol{k}) \pm \sqrt{\mathcal{M}(\boldsymbol{k})^2 + A_2^2 |\boldsymbol{k}|^2}. \tag{2.27}$$

Each of these is doubly degenerate due to the fact that $H(\boldsymbol{k})$ contains two uncoupled subspaces with equal eigenvalues. These subspaces have a well-defined *spin-orbit parity* (either $+1$ or -1), which is defined as the eigenvalue of $\sigma_z \otimes \tau_z$. Note that the operators τ_i are now understood to

act on the *hybridized* Bi$^+$ and Se$^-$ orbitals, so that nothing changes in practice. The dispersions are shown in Fig. 2.3, where we have also included the Dirac cone corresponding to the topological surface states. In particular, we note that the bulk bands of our effective model reproduce the expected band inversion around the Γ point.

The physical origin of the double degeneracy of each band is the presence of *both* time-reversal symmetry and inversion symmetry [88]. Under time reversal, $k \to -k$ and $\sigma \to -\sigma$ (where in this case σ represents the spin-orbit parity) and therefore $\epsilon_\sigma(k) = \epsilon_{-\sigma}(-k)$. On the other hand, inversion sends $k \to -k$ but leaves σ unchanged, and thus $\epsilon_\sigma(k) = \epsilon_\sigma(-k)$. The combination of both requirements leads to $\epsilon_\sigma(k) = \epsilon_{-\sigma}(k)$, and hence every band is doubly degenerate for arbitrary values of k.

We denote the eigenstates of $H(k)$ by $|\chi_k^\alpha\rangle$, with $\alpha \in \{c, v\} \cup \{+, -\}$. Here, c and v stand for the conduction and valence bands, respectively, and the sign corresponds to the spin-orbit parity. A convenient basis of eigenstates is given explicitly by

$$|\chi_k^{v,+}\rangle = \frac{1}{\sqrt{1+k^2 f(k)^2}} \begin{bmatrix} -f(k)k_- \\ 0 \\ 0 \\ 1 \end{bmatrix}, \quad |\chi_k^{v,-}\rangle = \frac{1}{\sqrt{1+k^2 f(k)^2}} \begin{bmatrix} 0 \\ 1 \\ -f(k)k_+ \\ 0 \end{bmatrix}, \quad (2.28\text{a})$$

$$|\chi_k^{c,+}\rangle = \frac{1}{\sqrt{1+k^2 f(k)^2}} \begin{bmatrix} 1 \\ 0 \\ 0 \\ f(k)k_+ \end{bmatrix}, \quad |\chi_k^{c,-}\rangle = \frac{1}{\sqrt{1+k^2 f(k)^2}} \begin{bmatrix} 0 \\ f(k)k_- \\ 1 \\ 0 \end{bmatrix}, \quad (2.28\text{b})$$

where we have defined

$$f(k) \equiv \frac{A_2}{\mathcal{M}(k) + \sqrt{\mathcal{M}(k)^2 + A_2^2 k^2}}, \quad \text{with } k \equiv |k|. \quad (2.29)$$

Note that the states as we have defined them are well-behaved for any value of k; in particular they are nonsingular at $k = 0$ since $M > 0$. On another note, their inner products satisfy

$$\langle \chi_k^{c,s} | \chi_q^{c,s} \rangle = \langle \chi_k^{c,-s} | \chi_q^{c,-s} \rangle^* = \langle \chi_k^{v,-s} | \chi_q^{v,-s} \rangle = \langle \chi_k^{v,s} | \chi_q^{v,s} \rangle^*, \quad (2.30\text{a})$$

$$\langle \chi_k^{c,s} | \chi_q^{v,s} \rangle = \langle \chi_k^{c,-s} | \chi_q^{v,-s} \rangle^* = -\langle \chi_k^{v,-s} | \chi_q^{c,-s} \rangle = -\langle \chi_k^{v,s} | \chi_q^{c,s} \rangle^*, \quad (2.30\text{b})$$

where $s \in \{+, -\}$.

2.5.2 Topological properties

Let us now show that the effective two-dimensional bulk Hamiltonian (2.26) preserves the topological properties of the original three-dimensional model. Note that, if we do not allow for perturbations that break time-reversal symmetry, then it belongs to the symmetry class AII. The $\mathbb{Z}_2$ topological invariant signals the existence of a quantum spin Hall phase.

To understand the topological nature of the Hamiltonian we begin by writing it as $H(k) = d(k) \cdot \Gamma$. Here we have defined $d(k) \equiv (A_2 k_x, A_2 k_y, \mathcal{M}(k))$ and $\Gamma \equiv (\sigma_x \otimes \tau_x, \sigma_y \otimes \tau_x, \tau_z)$, where the Kronecker products have been omitted. This reduces to two uncoupled Hamiltonians $H_\pm(k) = d_\pm(k) \cdot \sigma$ in the two spin-orbit subspaces, with $d_\pm(k) \equiv (A_2 k_x, A_2 k_y, \pm \mathcal{M}(k))$. We can therefore treat these Hamiltonians independently.

The topological properties of $H_\pm(k)$ can be traced back to the winding of the unit vector

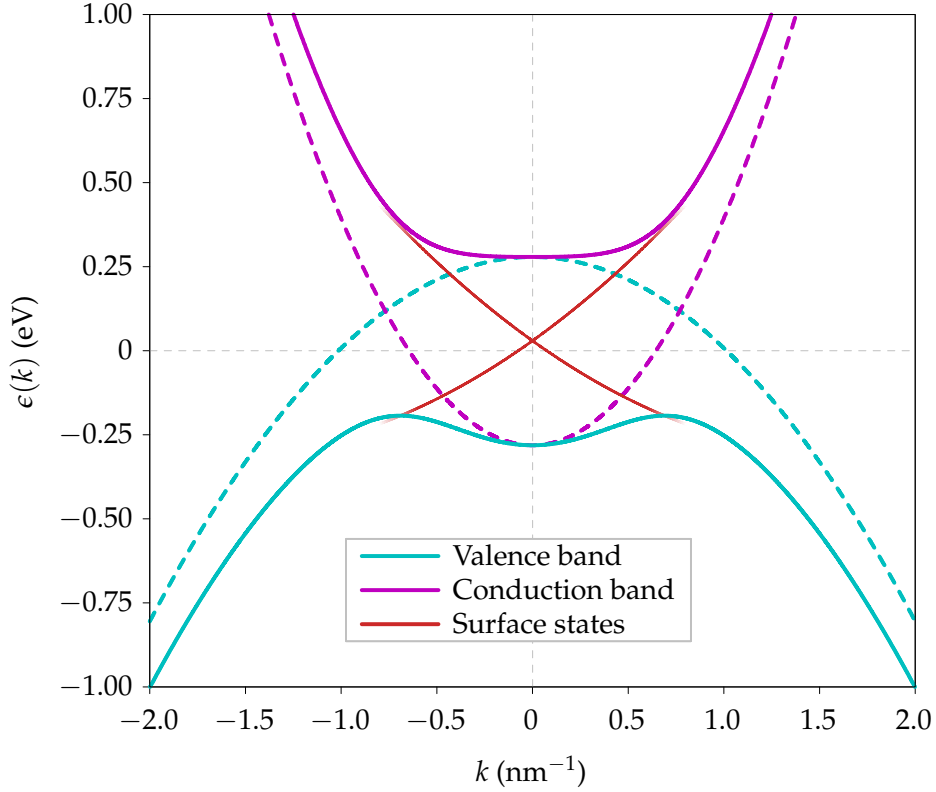$$n_\pm(k) = \frac{d_\pm(k)}{|d_\pm(k)|} \quad (2.31)$$

Figure 2.3. Band structure of two-dimensional Bi_2Se_3 around the Γ point of the Brillouin zone. Both the bulk conduction and valence bands are shown, as well as the gapless Dirac cones corresponding to the topologically protected surface states. The origin of energy corresponds to the Fermi level of undoped Bi_2Se_3. Note that the dispersions are isotropic in the employed $\boldsymbol{k} \cdot \boldsymbol{p}$ approximation, and that the bulk bands do not follow the usual parabolic dispersion relation. Instead, they display band inversion and an avoided crossing of the uncoupled states, which are represented with dashed lines. The conduction and valence band are each doubly degenerate due to the combined effect of inversion symmetry and time-reversal symmetry.

around the unit sphere as we move around the Brillouin zone. Note that this is well-defined everywhere in the Brillouin zone because the Hamiltonian is gapped for all momenta whenever $M \neq 0$ and $A_2 \neq 0$, which is assumed henceforth. We can compute the corresponding winding number with the expression introduced in Sec. 1.1.2, that is,

$$w_\pm = \frac{1}{4\pi} \int_{\text{BZ}} \mathrm{d}^2k \; \boldsymbol{n}_\pm \cdot \left(\frac{\partial \boldsymbol{n}_\pm}{\partial k_x} \times \frac{\partial \boldsymbol{n}_\pm}{\partial k_y} \right). \tag{2.32}$$

Computing the integrand gives

$$\boldsymbol{n}_\pm \cdot \left(\frac{\partial \boldsymbol{n}_\pm}{\partial k_x} \times \frac{\partial \boldsymbol{n}_\pm}{\partial k_y} \right) = \pm \frac{A_2^2(M + B_2 k^2)}{\left[(M - B_2 k^2)^2 + A_2^2 k^2 \right]^{3/2}}. \tag{2.33}$$

This must be integrated over our infinite Brillouin zone, yielding

$$w_\pm = \pm \frac{1}{2} (\operatorname{sgn} M + \operatorname{sgn} B_2). \tag{2.34}$$

There is a nontrivial winding whenever the signs of M and B_2 agree. In this regime, the Hamiltonian describes a QSH phase, and thus has a nontrivial $\mathbb{Z}_2$ topological invariant. When the signs of M and B_2 are opposite, the Hamiltonian describes a topologically trivial phase.

This agrees with the discussion of Refs. [35, 70], where the presented QSH Hamiltonians are (up to a unitary transformation) identical to our effective bulk Hamiltonian.

Even though the full Hamiltonian $H(\boldsymbol{k})$ is time-reversal invariant, the two blocks $H_{\pm}(\boldsymbol{k})$ individually break the time-reversal symmetry. Hence, they belong to the symmetry class AIII and can be classified according to a $\mathbb{Z}$ topological invariant. We have seen that this invariant corresponds to the Chern number of the eigenstates. Computing this quantity for the eigenstates in (2.28) gives

$$\mathcal{C}_{c,s} = -\mathcal{C}_{v,s} = \frac{s}{2}\left(\operatorname{sgn} M + \operatorname{sgn} B_2\right), \tag{2.35}$$

where $s \in \{+, -\}$. The detailed calculation of these Chern numbers can be found in App. 2.A, where we also give an expression for the corresponding Berry curvatures. Again, we are led to the conclusion that a topologically nontrivial phase arises when M and B_2 have the same sign.

Let us clarify, in passing, how the $\mathbb{Z}$ and $\mathbb{Z}_2$ topological invariants are related to each other. As discussed in Ref. [37], the bulk Chern number of each 2×2 block gives the number of chiral edge modes (counterclockwise minus clockwise) present at the boundary of a finite sample[1]. On the other hand, as explained in Sec. 1.2.4, the $\mathbb{Z}_2$ topological invariant dictates whether there is an even or an odd number of Dirac cones at the edge. In the nontrivial regime we have obtained Chern numbers of ± 1 for the filled bands of the block Hamiltonians, so there is always an extra chiral mode in each direction. The nontrivial $\mathbb{Z}_2$ invariant tells us that there is an odd number of Dirac cones. Since our Hamiltonian is two-dimensional, a Dirac "cone" in this case just corresponds to two oppositely moving modes that intersect linearly. These two modes are precisely the additional chiral states described by the unit Chern numbers.

---

[1]Note that here we are referring to the hypothetical boundary of a two-dimensional sample, and *not* to the top and bottom surfaces. Our nanosheets will always be assumed to be very large in the xy-directions, so the edge modes discussed in this paragraph will never play a role in the remainder of this work.

Appendices

2.A Berry curvature and Chern number of the eigenstates

In this section we compute the Chern number of the bulk eigenstates $|\chi_{\boldsymbol{k}}^{\alpha}\rangle$, which is well-defined whenever the energy gap does not close anywhere in the Brillouin zone. As shown in Fig. 2.3, this is the case for our model parameters of interest. We give the relevant steps for the state $|\chi_{\boldsymbol{k}}^{c,+}\rangle$; the rest can be computed in a similar fashion. Firstly, we have

$$\frac{\partial}{\partial k_x}|\chi_{\boldsymbol{k}}^{c,+}\rangle = \left[\frac{k_x}{k}\frac{\partial h}{\partial k_x}, \quad 0, \quad 0, \quad -\frac{k_x k_+}{k}\frac{\partial g}{\partial k} + g(k)\right]^{\mathsf{T}}, \tag{2.36a}$$

$$\frac{\partial}{\partial k_y}|\chi_{\boldsymbol{k}}^{c,+}\rangle = \left[\frac{k_y}{k}\frac{\partial h}{\partial k_x}, \quad 0, \quad 0, \quad \frac{k_y k_+}{k}\frac{\partial g}{\partial k} + ig(k)\right]^{\mathsf{T}}, \tag{2.36b}$$

where $g(k) \equiv f(k)[1 + k^2 f(k)^2]^{-1/2}$ and $h(k) \equiv [1 + k^2 f(k)^2]^{-1/2}$. Then the Berry curvature reads

$$\Omega_{xy}^{(c,+)}(\boldsymbol{k}) = 2\operatorname{Im}\langle\partial_{k_x}\chi_{\boldsymbol{k}}^{c,+}|\partial_{k_y}\chi_{\boldsymbol{k}}^{c,+}\rangle = 2g(k)^2\left[1 + k\frac{\partial}{\partial k}\log g(k)\right]. \tag{2.37}$$

By using the explicit form of $g(k)$, one finds

$$\Omega_{xy}^{(c,+)}(\boldsymbol{k}) = \frac{2kf(k)^2}{[1 + k^2 f(k)^2]^2}\frac{\partial}{\partial k}\log[kf(k)]. \tag{2.38}$$

For the rest of the states, one finds $\Omega_{xy}^{(c,-)} = -\Omega_{xy}^{(c,+)}$ and $\Omega_{xy}^{(v,s)} = -\Omega_{xy}^{(c,s)}$, with $s \in \{+, -\}$. The Chern number for each band must be computed from

$$\mathcal{C}_\alpha = \frac{1}{2\pi}\int d^2k\,\Omega_{xy}^{(\alpha)}(\boldsymbol{k}), \tag{2.39}$$

where the integral in this case runs over all space because in our continuous model the boundaries of the Brillouin zone have been sent to infinity. That is,

$$\mathcal{C}_{c,+} = \int_0^\infty dk\,\frac{2k^2 f(k)^2}{[1 + k^2 f(k)^2]^2}\frac{\partial}{\partial k}\log[kf(k)]. \tag{2.40}$$

The integral can be done by the change of variable $u(k) = 1 + k^2 f(k)^2$, because

$$\mathrm{d}u = 2kf(k)\frac{\partial}{\partial k}[kf(k)] = 2k^2 f(k)^2 \frac{\partial}{\partial k}\log[kf(k)]. \tag{2.41}$$

Then we simply find

$$\mathcal{C}_{c,+} = \int_{u(0)}^{u(\infty)} \frac{\mathrm{d}u}{u^2}. \tag{2.42}$$

Everything we have done so far is completely general and holds for arbitrary values of the model parameters, provided there is always a nonzero gap between the valence and conduction bands. The Chern number ultimately only depends on the values of $u(0)$ and $u(\infty)$. These limits must be taken with care. We recall that

$$f(k) = \frac{A_2}{\mathcal{M}(k) + \sqrt{\mathcal{M}(k)^2 + A_2^2 k^2}}, \tag{2.43}$$

where $\mathcal{M}(k) = M - B_2 k^2$, and note that the sign of A_2 is unimportant because $u(k)$ only depends on $f(k)^2$. The only requirement is $A_2 \neq 0$ in order to be in the insulating regime. One now may verify the following table:

	$M > 0$	$M < 0$
$B_2 > 0$	$u(0) = 1, u(\infty) = \infty$	$u(0) = \infty, u(\infty) = \infty$
$B_2 < 0$	$u(0) = 1, u(\infty) = 1$	$u(0) = \infty, u(\infty) = 1$

It immediately follows that the Chern number vanishes when $MB_2 < 0$. By contrast, $\mathcal{C}_{c,+} = +1$ when $M > 0$, $B_2 > 0$, and $\mathcal{C}_{c,+} = -1$ when $M < 0$, $B_2 < 0$. In our model, the former is the relevant case. A similar procedure yields $\mathcal{C}_{c,-} = -\mathcal{C}_{c,+}$ and $\mathcal{C}_{v,s} = -\mathcal{C}_{c,s}$. We can summarize the results as

$$\mathcal{C}_{c,s} = -\mathcal{C}_{v,s} = \frac{s}{2}(\operatorname{sgn} M + \operatorname{sgn} B_2). \tag{2.44}$$

Note that $\sum_s \mathcal{C}_{v,s} = 0$ as expected from the fact that the total Hamiltonian has time-reversal symmetry and inversion symmetry, so that the sum of Berry curvatures of all occupied bands vanishes [30]. Similarly, the result $\mathcal{C}_{v,s} = -\mathcal{C}_{c,s}$ follows from the fact that the total Berry curvature vanishes separately for each of the irreducible blocks that make up the total Hamiltonian [89, Ch. 4]. We see that the nonzero Chern numbers arise whenever $MB_2 > 0$. This is precisely the case for which there exists a band inversion. In this regime the insulator is in a topologically nontrivial phase, in agreement with Ref. [35]. By continuously modifying one of the parameters, either M or B_2, the gap eventually closes, and it reopens with the insulator in a trivial state. This corresponds to a topological phase transition.

3

Excitons in Bi$_2$Se$_3$

Now that we have obtained a low-energy model describing the physics of noninteracting particles, we will begin our treatment of excitons by adding the electromagnetic interaction between electrons and holes. We will first solve for the exciton dispersions and wave functions for arbitrary values of the total quasiparticle momentum. The features of the exciton spectrum in topological Bi$_2$Se$_3$ will be explained analytically and compared with the situation in a trivial insulator. We will also check our first-band approximation introduced in the previous chapter and verify that our approach is self-consistent. Finally, an effective low-energy model for excitons will be derived and used to elucidate the topological properties of the excitonic band structure.

3.1   Electrons and holes

So far we have cast the eigenstates in terms of conduction and valence states. However, in many applications it will be useful to work in the physically equivalent electron-hole picture (note that here the term *electron* refers exclusively to conduction electrons). Mathematically, the transformation of one picture to the other is simple, yet slightly subtle for holes. To understand it precisely it is best to consider the problem in the second-quantized formalism. The field operators in the conduction-valence picture are written as [90, Ch. 6]

$$\hat{\psi}_c^{i,\sigma}(\boldsymbol{x}) = \sum_{\boldsymbol{k},s} \hat{c}_{\boldsymbol{k},s}\, \chi_{\boldsymbol{k};c,s}^{i,\sigma}(\boldsymbol{x})\,, \tag{3.1a}$$

$$\hat{\psi}_v^{i,\sigma}(\boldsymbol{x}) = \sum_{\boldsymbol{k},t} \hat{v}_{\boldsymbol{k},t}\, \chi_{\boldsymbol{k};v,t}^{i,\sigma}(\boldsymbol{x})\,, \tag{3.1b}$$

where $\hat{c}$ and $\hat{v}$ are the annihilation operators of conduction and valence electrons, respectively. The indices $i \in \{\mathrm{Bi}^+, \mathrm{Se}^+\}$ and $\sigma \in \{\uparrow, \downarrow\}$ denote our (hybridized) orbital and spin states, respectively. The single-particle wave functions are defined as

$$\chi_{\boldsymbol{k};\alpha}^{i,\sigma}(\boldsymbol{x}) \equiv \langle \boldsymbol{x}; i, \sigma | \boldsymbol{k}; \alpha \rangle = \frac{\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}}{\sqrt{V}} \langle i, \sigma | \chi_{\boldsymbol{k}}^{\alpha} \rangle\,, \tag{3.2}$$

41

where the states $|\chi_k^\alpha\rangle$ are given in Eqs. (2.28). On the other hand, we may write down the field operators in the electron-hole picture as

$$\hat{\psi}_e^{i,\sigma}(x) = \sum_{k,s} \hat{e}_{k,s} \chi_{k;e,s}^{i,\sigma}(x), \tag{3.3a}$$

$$\hat{\psi}_h^{i,\sigma}(x) = \sum_{k,t} \hat{h}_{k,t} \chi_{k;h,t}^{i,\sigma}(x), \tag{3.3b}$$

where $\hat{e}$ and $\hat{h}$ are the annihilation operators of (conduction) electrons and holes, respectively. The key point now is that creating a hole at position x in the spin-orbit state (i, σ) is equivalent to destroying a valence electron at the same position and state. Hence, $\psi_v^{i,\sigma}(x)$ must be identified with $\hat{\psi}_h^{i,\sigma\dagger}(x)$. Furthermore, the hole that is created must have opposite momentum to that of the destroyed electron, so $\hat{v}_{k,t}$ is identified with $\hat{h}^\dagger_{-k,t}$. Then, it follows that the valence-electron and hole wave functions are related through

$$\langle x; i, \sigma | k; h, t \rangle = \langle x; i, \sigma | -k; v, t \rangle^* = \frac{e^{ik \cdot x}}{\sqrt{V}} \langle \chi^\alpha_{-k} | i, \sigma \rangle. \tag{3.4}$$

On the other hand, the electron states remain unchanged with respect to the conduction-valence picture.

3.2 Exciton states

Consider an electron and a hole at positions r_e and r_h and with momenta $k_e \equiv Q/2 + k$ and $k_h \equiv Q/2 - k$, respectively, bound into an exciton with total momentum Q. Here, k is a relative momentum between the two particles. The real-space coordinates conjugate to Q and k are denoted by R and r, respectively. The following relations are satisfied:

$$Q = k_e + k_h, \qquad k = \frac{k_e - k_h}{2}, \tag{3.5a}$$

$$R = \frac{r_e + r_h}{2}, \qquad r = r_e - r_h. \tag{3.5b}$$

Note that we make no attempt to use the center-of-mass coordinate system because the electron and hole masses are not well-defined due to the band inversion. The equations above are acceptable because they preserve the canonical commutation relations in the operator language, $[r_i, k_j] = [R_i, Q_j] = i\delta_{ij}$.

Since the two particles interact with each other, their relative momentum k will not be conserved and only Q will be a constant of the motion. A general exciton state is then a superposition of states with all possible k. Let us assume for the moment that the electron and the hole have well-defined spin-orbit parities s and t, respectively. A generic exciton state with total momentum Q can then be written as

$$|Q; s, t\rangle = \frac{1}{\sqrt{V}} \sum_k \Phi_{Q;s,t}(k) \hat{e}^\dagger_{Q/2+k,s} \hat{h}^\dagger_{Q/2-k,t} |G\rangle. \tag{3.6}$$

Here, $|G\rangle$ denotes the neutral semiconductor ground state, which is killed by the annihilation operators of both holes and conduction electrons. Also, $\Phi_{Q;s,t}(k)$ is the relative electron-hole wave function, whose determination is one of the objectives of this chapter. As we will see in the next section, the interaction can also couple states with different spin-orbit parities, so the true eigenstates of the full interacting problem will be linear combinations of the different states $|Q; s, t\rangle$.

We further note that, in reality, $|\boldsymbol{Q}; s, t\rangle$ stands for an entire *family* of states that need to be labeled by additional quantum numbers, since (in analogy to the hydrogen problem) there will in general be many different solutions for the excitonic wave functions. For the moment, however, we just leave this additional dependence implicit; later we shall introduce a notation to label the individual states within these families.

3.3 Effective potential for the electron-hole interaction

Consider an exciton with total momentum $\boldsymbol{Q}$ and relative momentum $\boldsymbol{k}'$. The band structure of Bi_2Se_3 modifies the interaction, V, between the electron and the hole. These may scatter into a new state with relative momentum $\boldsymbol{k}$. Note, however, that the total momentum $\boldsymbol{Q}$ is conserved. To lowest order, this effect is captured by the following direct and exchange Feynman diagrams:
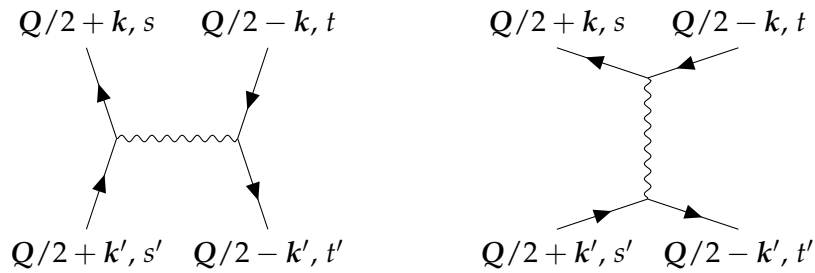
Evaluating them (cf. App. 3.A for a detailed calculation) gives the matrix elements for the effective interaction. The contribution to the scattering amplitude of the direct and exchange diagrams reads, respectively,

$$V^{\mathrm{D}}_{s,t;s',t'}(\boldsymbol{Q};\boldsymbol{k},\boldsymbol{k}') = \delta_{ss'}\delta_{tt'}V(\boldsymbol{k}-\boldsymbol{k}')\langle\chi^{c,s}_{\boldsymbol{Q}/2+\boldsymbol{k}}|\chi^{c,s}_{\boldsymbol{Q}/2+\boldsymbol{k}'}\rangle\langle\chi^{v,t}_{-\boldsymbol{Q}/2+\boldsymbol{k}'}|\chi^{v,t}_{-\boldsymbol{Q}/2+\boldsymbol{k}}\rangle,\tag{3.7a}$$

$$V^{\mathrm{X}}_{s,t;s',t'}(\boldsymbol{Q};\boldsymbol{k},\boldsymbol{k}') = \delta_{st}\delta_{s't'}V(\boldsymbol{Q})\langle\chi^{c,s}_{\boldsymbol{Q}/2+\boldsymbol{k}}|\chi^{v,s}_{-\boldsymbol{Q}/2+\boldsymbol{k}}\rangle\langle\chi^{v,s'}_{-\boldsymbol{Q}/2+\boldsymbol{k}'}|\chi^{c,s'}_{\boldsymbol{Q}/2+\boldsymbol{k}'}\rangle.\tag{3.7b}$$

Here, $V(\boldsymbol{q})$ stands for the bare electrostatic momentum-space interaction between electrons and holes within the nanosheet. In the low-momentum regime $q \to 0$ we always expect the potential to acquire the usual two-dimensional Coulomb form, proportional to q^{-1}. One may verify that the inner products in Eq. (3.7b) tend to zero linearly with Q, and thus the exchange interaction always vanishes at $\boldsymbol{Q} = \boldsymbol{0}$.

Which precise form to choose for the interaction potential for larger momenta is a quite nontrivial question. Even though the topology of the problem will be unaffected by the particular choice of V, the numerical values of, e.g., exciton binding energies and radii may still be greatly affected by it. For this reason, the next section presents a detailed discussion of typically used approximations for the electron-hole interaction in semiconductor nanosheets.

3.3.1 Discussion of the interaction potential

Consider a thin semiconducting layer of thickness ℓ occupying the region $-\ell/2 \leq z \leq \ell/2$. The film possesses a bulk dielectric constant ε_d in the direction parallel to the substrate, and the surrounding environment has a dielectric constant $\varepsilon_s < \varepsilon_d$. Then the electrostatic interaction energy between two charges e and $-e$ located at points $(\boldsymbol{r}, z)$ and $(0, z')$ (with $\boldsymbol{r}$ the in-plane coordinates and $z \geq z'$) is given by the *Rytova-Keldysh* (RK) potential [91],

$$V^{\mathrm{RK}}(\boldsymbol{r};z,z') = -\frac{e^2}{\varepsilon_0\varepsilon_d}\int\frac{\mathrm{d}^2k}{(2\pi)^2}\,\mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{r}}\,\frac{\cosh\left[k(\ell/2-z)+\eta\right]\cosh\left[k(\ell/2+z')+\eta\right]}{k\sinh\left[k\ell+2\eta\right]},\tag{3.8}$$

where

$$\eta = \frac{1}{2} \log\left(\frac{\varepsilon_d + \varepsilon_s}{\varepsilon_d - \varepsilon_s}\right). \tag{3.9}$$

When $\varepsilon_s \ll \varepsilon_d$ and $\ell \ll r$, the Rytova-Keldysh potential reduces to the so-called *Struve-Neumann* (SN) potential, which reads

$$V^{\mathrm{SN}}(r) = -\frac{e^2}{4\varepsilon_0 \varepsilon_d \ell}\left[\mathbf{H}_0\left(\frac{2\varepsilon_s}{\varepsilon_d}\frac{r}{\ell}\right) - Y_0\left(\frac{2\varepsilon_s}{\varepsilon_d}\frac{r}{\ell}\right)\right]. \tag{3.10}$$

Here, Y_0 is the Bessel function of the second kind (also called the Neumann function) and $\mathbf{H}_0$ is the Struve function. Note that, rather confusingly, the Struve-Neumann potential is sometimes also called the Rytova-Keldysh potential in the literature [92, 93]. For even larger distances, $r \gg (\varepsilon_d/2\varepsilon_s)\ell$, this potential further reduces to the Coulomb potential with the permittivity of the surrounding medium,

$$V^{\mathrm{C}}(r) = -\frac{e^2}{4\pi\varepsilon_0\varepsilon_s r}. \tag{3.11}$$

Physically, this is to say that in this large-thickness limit the electric field lines connecting both charges lie mainly outside of the nanosheet.

In momentum space, the Struve-Neumann potential is

$$V^{\mathrm{SN}}(k) = -\frac{e^2}{2\varepsilon_0\varepsilon_s}\frac{1}{k\left[1 + (\varepsilon_d/2\varepsilon_s)\ell k\right]} \tag{3.12}$$

and the Coulomb potential is

$$V^{\mathrm{C}}(k) = -\frac{e^2}{2\varepsilon_0\varepsilon_s k}. \tag{3.13}$$

Note that these potentials arise from purely classical considerations. Thus, they accurately capture the behavior of the involved particles at long distances. At short distances, however, the effects of quantum confinement become important and the resulting effective potential may significantly differ from that given by Eq. (3.8). These effects will be irrelevant if the obtained exciton radii are indeed within the regime of validity of the SN potential.

We have performed the subsequent numerical calculations with both the Struve-Neumann potential and the Coulomb potential. We have chosen the relative permittivity $\varepsilon_s = 6$, which is a typical low value for the environment [94]. The bulk dielectric constant has been set to $\varepsilon_d = 28$ in accord with recent first-principles studies of rhombohedral Bi_2Se_3 in the near-infrared region [95, 96]. Furthermore, we use $e^2/\varepsilon_0 = 18.0951\,\mathrm{eV\,nm}$ everywhere. Once the excitonic wave functions corresponding to these potentials have been determined, the characteristic exciton size can be found for each bound state by calculating the mean squared diameter, $a_0 \equiv \sqrt{\langle r^2 \rangle}$. This average distance between the electron and the hole can then be compared with the length scales corresponding to the regimes of validity of both potentials. In this manner, the accuracy of our calculations may easily be checked *a posteriori*.

Our findings, reported in the following sections, indicate that the long-distance Coulomb potential is not a good approximation, as the exciton diameters of the lowest-lying states are actually smaller than the chosen thickness of the nanosheet. Nevertheless, the tightly bound exciton states resulting from the use of the bare Coulomb interaction allow us to clearly visualize the effects of the topology, as discussed in Sec 3.8. Furthermore, the results obtained with the use of the Struve-Neumann potential also *a priori* seem to indicate that this potential is again not a satisfactory approximation, as the exciton radii are comparable or just slightly larger than the slab thickness. Nevertheless, the validity of the calculations is restored by a coincidence reported in Ref. [97], which provides a detailed study on the effects of quantum

confinement in semiconductor nanosheets. It is found that, both for small *and* large momenta, the classical Struve-Neumann interaction very closely resembles the Rytova-Keldysh potential with quantum-confinement effects included. In other words, even though the bare RK and SN potentials start to differ significantly in the regime $k\ell > 1$, corrections introduced to the RK interaction by quantum confinement make it similar to the SN interaction even for $k\ell > 1$, as can be appreciated in Fig. 7 in the article. We stress that this is a coincidence with no meaningful physical interpretation, but it makes our predictions valid, at least in the context of the first-band approximation of Sec. 2.4, as our lowest conduction and valence probability densities resemble that of the infinite-well ground state considered in the paper. Indeed, we have checked that for $Q = 0$ the differences between exciton binding energies resulting from using the SN and the quantum-confined RK potentials are of the order of 1 meV, whereas the binding energies themselves are of the order of 15–25 meV. In any case, the soundness of the results obtained with the SN interaction must ultimately be checked by comparing the exciton binding energies with the splitting of the bulk subbands, as explained in Sec. 2.4.1.

Another important question is that of screening, by which the bulk interaction potential may be modified due to the creation and annihilation of virtual electron-hole pairs. We have opted to neglect this effect in the numerical calculations of this chapter, which can be partially motivated as follows. We expect the screening to be caused mainly by the virtual excitation of surface particles, whose dispersion forms a gapless Dirac cone. By contrast, the contribution of virtual bulk electron-hole pairs should be heavily suppressed in view that the bulk of Bi_2Se_3 is insulating with an energy gap E_g[1]. However, the so-called *Thomas-Fermi momentum*, corresponding to the inverse of the characteristic screening length, should be proportional to the density of states at the Fermi level for low frequencies [98, Ch. 2]. Since this vanishes at the Dirac point, we expect no important contribution to the screening arising from the direct diagram in Sec. 3.3 if the chemical potential is tuned appropriately. Assuming that this is the case, it is correct to disregard the screening of the direct interaction in first approximation. On the other hand, however, the dominant frequency contribution to the exchange diagram takes place at the total exciton energy. In this case we cannot expect the low-frequency limit to hold and there will be in principle a screening contribution, as interband (valence-conduction) transitions of surface electrons are possible even when the chemical potential is located at the Dirac point. Nevertheless, the exchange interaction vanishes for $Q \to 0$, and thus this effect is not important around the Γ point. For larger values of Q, we have to resort to an *a posteriori* justification of the perturbative nature of the corresponding screening effects. In Ch. 4 we shall calculate the total energy shift in perturbation theory and discuss the validity of this approximation.

3.4 The exciton eigenvalue problem

We are now interested in finding the exciton states that diagonalize the Coulomb problem with both direct and exchange interactions present. The most general exciton state with total momentum Q is written as

$$\left|\Psi_Q\right\rangle = \frac{1}{\sqrt{V}} \sum_{k,s,t} \Phi_{Q;s,t}(k)\, \hat{e}^\dagger_{Q/2+k,s}\, \hat{h}^\dagger_{Q/2-k,t} \left|G\right\rangle. \tag{3.14}$$

Note that the coefficients of the linear combination of individual states $|Q; s, t\rangle$ have been absorbed into the wave functions. The expected value of the full two-body Hamiltonian in

---

[1]Quantitatively, the probability of excitation of bulk electron-hole pairs will be a factor of $\exp(E_g/k_B T)$ smaller than that of surface pairs. This only becomes of order unity around $k_B T \sim E_g \approx 0.22$ eV for bismuth selenide, corresponding to $T \sim 2500$ K, way above our temperature scales of interest.

the exciton state $|\Psi_Q\rangle$ is

$$\langle \Psi_Q | \hat{\mathcal{H}} | \Psi_Q \rangle = \frac{1}{V} \sum_{k,k'} \sum_{s,s'} \sum_{t,t'} \Phi^*_{Q;s,t}(k) \langle Q, k; s, t | H | Q, k'; s', t' \rangle \Phi_{Q;s',t'}(k'), \qquad (3.15)$$

where the Hamiltonian matrix elements read

$$\langle Q, k; s, t | H | Q, k'; s', t' \rangle = \left[\epsilon_c(Q/2 + k) - \epsilon_v(Q/2 - k) \right] \delta_{k,k'} \delta_{s,s'} \delta_{t,t'}$$
$$+ \frac{1}{V} \left[V^D_{s,t;s',t'}(Q; k, k') - V^X_{s,t;s',t'}(Q; k, k') \right]. \qquad (3.16)$$

Note that the minus sign in front of ϵ_v comes from the fact that the energy of a hole is opposite of that of the corresponding electron[2]. Again, we note that $|\Psi_Q\rangle$ in Eq. (3.14) must be understood as a whole family of states that solve the exciton problem, which need to be labeled with additional quantum numbers that we have omitted for the sake of brevity.

The exciton eigenvalue problem is now obtained by minimizing the following energy functional with respect to the wave functions $\Phi_{Q;s,t}$, which are taken as the variational parameters:

$$\mathcal{F}[\Phi] = \langle \Psi_Q | \hat{\mathcal{H}} | \Psi_Q \rangle - \varepsilon_Q \langle \Psi_Q | \Psi_Q \rangle. \qquad (3.17)$$

Here we have added a Lagrange multiplier in order to account for the normalization condition $\langle \Psi_Q | \Psi_Q \rangle = \frac{1}{V} \sum_{k,s,t} |\Phi_{Q;s,t}(k)|^2 = $ constant. This leads to a *Bethe-Salpeter equation* (BSE):

$$\sum_{k',s',t'} \langle Q, k; s, t | H | Q, k'; s', t' \rangle \Phi_{Q;s',t'}(k') = \varepsilon_Q \Phi_{Q;s,t}(k). \qquad (3.18)$$

In general, our exciton eigenstates $|\Psi_Q\rangle$ are linear combinations of the basis states $|Q; s, t\rangle$. The problem can be greatly simplified by noting that the two states with $s \neq t$ are not coupled by the exchange interaction, nor are they coupled to the subspace spanned by the states with $s = t$. Hence, $|Q; +, -\rangle$ and $|Q; -, +\rangle$ are eigenstates of the Hamiltonian, and the BSE can be solved in the subspaces spanned by these individual states. Furthermore, the properties in Eq. (2.30) imply that

$$V^{D/X}_{s,t;s',t'}(Q; k, k') = V^{D/X}_{-s,-t;-s',-t'}(Q; k, k')^*, \qquad (3.19)$$

so the two states with $s \neq t$ are degenerate in energy for any Q and $\Phi_{Q;+-}(k)$ can be identified with $\Phi^*_{Q;-+}(k)$. On the other hand, the states $|Q; +, +\rangle$ and $|Q; -, -\rangle$ *are* coupled by the exchange interaction, so the BSE must be solved in this larger subspace. Nevertheless, for small Q the inner products in Eq. (3.7b) tend to zero linearly with Q, so $|0; +, +\rangle$ and $|0; -, -\rangle$ are degenerate eigenstates of the Hamiltonian, again with $\Phi_{0;++}(k) = \Phi^*_{0;--}(k)$ up to an arbitrary phase.

To numerically solve the problem, the wave functions $\Phi_{Q;s,t}(k)$ are written as a Fourier-series expansion,

$$\Phi_{Q;s,t}(k) = \sum_m \Phi^{(m)}_{Q;s,t}(k) \, e^{im\phi_k}. \qquad (3.20)$$

Here, ϕ_k is the angle between k and the x-axis, which will simply be denoted by ϕ when there is no risk of confusion. The full exciton problem then reads

$$\sum_{m'} \left[\mathcal{K}^{(m,m')}(Q; k) \Phi^{(m')}_{Q;s,t}(k) + \sum_{s',t'} \int_0^\infty \frac{dk'}{2\pi} k' \mathcal{V}^{(m,m')}_{s,t;s',t'}(Q; k, k') \Phi^{(m')}_{Q;s',t'}(k') \right] = \varepsilon_Q \Phi^{(m)}_{Q;s,t}(k), \quad (3.21)$$

---

[2]Strictly speaking, this is only the case when the chemical potential coincides with the zero of energy. Since in the previous section we have assumed a shift in the chemical potential so that it coincides with the Dirac point, we take this as our new origin of energy.

where we have defined

$$
\mathcal{K}^{(m,m')}(\boldsymbol{Q};k) = \int_0^{2\pi} \frac{\mathrm{d}\phi}{2\pi}\, \mathrm{e}^{-\mathrm{i}(m-m')\phi} \left[\epsilon_c(\boldsymbol{Q}/2 + \boldsymbol{k}_\phi) - \epsilon_v(\boldsymbol{Q}/2 - \boldsymbol{k}_\phi) \right], \tag{3.22a}
$$

$$
\mathcal{V}_{s,t;s',t'}^{(m,m')}(\boldsymbol{Q};k,k') = \int_0^{2\pi} \frac{\mathrm{d}\phi}{2\pi} \int_0^{2\pi} \frac{\mathrm{d}\phi'}{2\pi}\, \mathrm{e}^{-\mathrm{i}m\phi}\mathrm{e}^{\mathrm{i}m'\phi'} \left(V^{\mathrm{D}} - V^{\mathrm{X}} \right)_{s,t;s',t'}(\boldsymbol{Q};\boldsymbol{k}_\phi,\boldsymbol{k}'_{\phi'}), \tag{3.22b}
$$

with $\boldsymbol{k}_\phi \equiv (k\cos\phi, k\sin\phi)$ and $\boldsymbol{k}'_{\phi'} \equiv (k'\cos\phi', k'\sin\phi')$.

3.4.1 The case $\phi_Q = 0$

When solving the problem numerically we choose $\boldsymbol{Q}$ along the x-direction, i.e., $\boldsymbol{Q} = Q\hat{x}$. When $\boldsymbol{Q}$ forms a nonzero angle ϕ_Q with the x-axis the problem can be studied analytically with knowledge about the case with $\phi_Q = 0$, as will be done in the next section. For $\phi_Q = 0$, the potentials satisfy the property

$$
V_{s,t;s',t'}^{\mathrm{D/X}}(Q\hat{x};\boldsymbol{k}_\phi,\boldsymbol{k}'_{\phi'}) = V_{s,t;s',t'}^{\mathrm{D/X}}(Q\hat{x};\boldsymbol{k}_{-\phi},\boldsymbol{k}'_{-\phi'})^*, \tag{3.23}
$$

which implies that the matrix elements $\mathcal{V}_{s,t;s',t'}^{(m,m')}(Q\hat{x};k,k')$ are all real (cf. App. 3.C). The same holds for $\mathcal{K}^{(m,m')}(Q\hat{x};k)$, as for zero ϕ_Q the imaginary part of the integrand in (3.65a) is odd in ϕ. It follows that the Hamiltonian matrix in (3.64) is real and Hermitian. Thus, the Fourier coefficients $\Phi_{Q\hat{x};s,t}^{(m)}(k)$ that diagonalize the problem can all be simultaneously chosen as real[3].

3.4.2 The case $\phi_Q \neq 0$

The topological properties of the excitonic states can only be determined with knowledge about the states on the entire $\boldsymbol{Q}$-space, since we need knowledge about the states for arbitrary $\boldsymbol{Q}$ in order to compute the Berry connection and curvature. This means that solving the problem for $\boldsymbol{Q} = Q\hat{x}$ is not enough to discuss the topology and we need to study the general problem in which $\boldsymbol{Q}$ forms a nonzero angle ϕ_Q with the x-axis. A careful analysis is carried out in App. 3.D, where we also give relations between the wave functions at zero and nonzero ϕ_Q. The most important conclusion is that the excitonic eigenstates split into two doublets, $|\boldsymbol{Q};0_\pm\rangle$ and $|\boldsymbol{Q};2_\pm\rangle$, given by

$$
|\boldsymbol{Q};0_\pm\rangle = \frac{1}{\sqrt{2}}\Big(|\boldsymbol{Q};+,-\rangle \pm |\boldsymbol{Q};-,+\rangle \Big), \tag{3.24a}
$$

$$
|\boldsymbol{Q};2_\pm\rangle = \frac{1}{\sqrt{2}}\Big(\mathrm{e}^{-\mathrm{i}\phi_Q}|\boldsymbol{Q};+,+\rangle \pm \mathrm{e}^{\mathrm{i}\phi_Q}|\boldsymbol{Q};-,-\rangle \Big). \tag{3.24b}
$$

In these, the convention is that the wave functions contained in $|\boldsymbol{Q};+,-\rangle$ and $|\boldsymbol{Q};-,+\rangle$ are complex conjugates of each other, and similarly for those in $|\boldsymbol{Q};+,+\rangle$ and $|\boldsymbol{Q};-,-\rangle$. Due to the combined phase factors $\mathrm{e}^{\pm\mathrm{i}\phi_Q}$, the members of the doublet $|\boldsymbol{Q};2_\pm\rangle$ have chirality two. We can define a spin-orbit-parity pseudospin operator $\boldsymbol{\sigma} \equiv (\sigma_x, \sigma_y, \sigma_z)$ that acts on the space spanned by $|\boldsymbol{Q};+,+\rangle$ and $|\boldsymbol{Q};-,-\rangle$. Its expected value for the states $|\boldsymbol{Q};2_\pm\rangle$ is

$$
\langle\boldsymbol{\sigma}\rangle_\pm = \pm\cos 2\phi_Q\,\hat{x} \pm \sin 2\phi_Q\,\hat{y}. \tag{3.25}
$$

The corresponding pseudospin texture is sketched in Fig. 3.1. We see that it has a winding number of 2 around $\boldsymbol{Q} = \boldsymbol{0}$, analogously to the valley pseudospin in Ref. [100].

---

[3]Actually, the integration measure slightly modifies the Hermiticity property to $H^{(m)}(k,k') = (k'/k)H^{(m)}(k',k)^*$ [99], but this does not invalidate the conclusion because it is merely an artifact of the continuum limit.

Note that the wave functions in $|Q; 2_+\rangle$ are in general different from those in $|Q; 2_-\rangle$. This is to say that the states $|Q; +, +\rangle$ and $|Q; -, -\rangle$ are not actually well-defined on their own in the presence of an exchange interaction, which couples them and modifies them. However, the expressions above still make sense for small Q, and more generally if we take into account that the wave functions depend on the eigenstate itself.
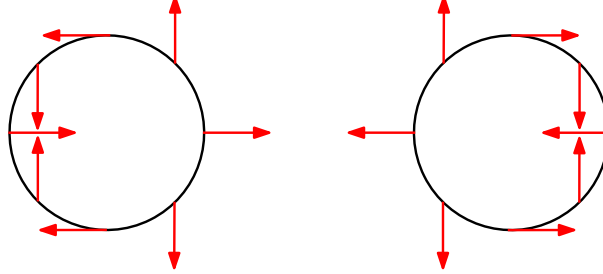

Figure 3.1. Orientation of the spin-orbit pseudospin vector σ for the states $|Q; 2_+\rangle$ (left) and $|Q; 2_-\rangle$ (right) as the polar angle ϕ_Q varies from 0 to 2π. The pseudospin winds twice around itself in this process.

3.5   Solution of the Bethe-Salpeter equation

In this section we discuss the solutions to the BSE (3.64). The numerical method we have employed is based on Ref. [99]; a brief overview is given in App. 3.B. We have only considered angular momenta $|m| \leq 3$ after verifying that higher-m corrections do not play any significant role for the low-lying states.

3.5.1   Labeling of the eigenstates

Before discussing the results, let us introduce a notation to label the different individual states within the families $|Q; 2_\pm\rangle$ and $|Q; 0_\pm\rangle$ belonging to the subspaces with $s = t$ and $s \neq t$, respectively. It will be enough to specify two additional quantum numbers in order to identify them unambiguously: a principal quantum number $n \in \{0, 1, 2, \ldots\}$, and the azimuthal angular momentum m at $Q = 0$, which can be chosen to satisfy $m \in \{-n, \ldots, 0, \ldots, n\}$. In this way, the states are labeled in the same way as the orbitals of the 2D hydrogen problem [101]. The number m is well-defined due to the rotational invariance of the problem *at zero momentum*. Away from $Q = 0$ the states are labeled in the same way by connecting them to their zero-momentum counterparts. This can be done unambiguously for the subspaces with $s = t$ and $s \neq t$, as the kinetic matrix couples all angular momenta and thus there exist no level crossings within each subspace. We can now label individual states as $|Q; 2_\pm; n, m\rangle$ and $|Q; 0_\pm; n, m\rangle$.

In practice, we first identify m from the numerically obtained momentum-space wave functions and *then* assign a value of n that conforms to the rule above[4]. The number of nodes of the radial wave function is $n - |m| + (1 - \delta_{m,0})$ in the case of the hydrogen atom. Because we are only interested in the low-lying states, we have not obtained enough excited states to verify if this is also the case for topologically nontrivial Bi_2Se_3. However, we find that the ordering is completely different from that of the hydrogen spectrum, which goes as $-\left(n + \frac{1}{2}\right)^{-2}$. For instance, the 1$s$ state has a higher energy than the 1p, 2d and 3f states.

---

[4]It is worth mentioning that some care must be taken when doing this. The gauge freedom $|\chi_q^\alpha\rangle \to e^{i\zeta_\alpha(q)}|\chi_q^\alpha\rangle$ of the single-particle eigenstates shifts the interaction in the mth channel to the $(m+w)$th channel in the transformed basis, where w is the winding number of $\zeta_\alpha(q)$ [102]. Therefore, one ought to be careful when labeling exciton states with their azimuthal angular momentum. The ambiguity disappears when demanding that the single-particle eigenstates be nonsingular at $q = 0$, which is indeed the case for the states of Eq. (2.28) [103].

3.5.2 Exciton energy spectrum

Fig. 3.2 shows the dispersions of the excitonic eigenstates arising from the BSE. The two plots correspond to the results obtained by using the Coulomb and Struve-Neumann potentials. The different colors label the absolute value of the azimuthal angular momentum m at $Q = 0$. The solid, dashed and dotted lines correspond to states pertaining to the families $|Q; 2_+\rangle$, $|Q; 2_-\rangle$, and $|Q; 0_\pm\rangle$, respectively.

The upper magenta region in both figures is the electron-hole continuum, that is, the region without bound electron-hole states. It can be obtained numerically by completely switching off the interaction between particles. For fixed Q, it is given analytically by the minimum value of the energy gap

$$\Delta_Q(k) \equiv \epsilon_c(Q/2 + k) - \epsilon_v(Q/2 - k). \tag{3.26}$$

We have obtained it through both methods. In the figures, the magenta solid line corresponds to minimizing Δ_Q and the magenta dashed line is the continuum obtained numerically by keeping $|m| \leq 3$. We have checked that the dashed line converges to the analytical curve upon the inclusion of higher angular momenta.

Turning to the figures, the first thing to notice is the difference in behavior between the ground states of the $|Q; 2_\pm\rangle$ and $|Q; 0_\pm\rangle$ families (i.e., the states $|Q; 2_\pm; 0, 0\rangle$ and $|Q; 0_\pm; 0, 0\rangle$) around $Q = 0$, as the former one splits off into one linear mode and one quadratic mode. This may be analytically understood by expanding the coupling term $V^X_{++;--}(Q; k, k')$ in the effective potential around this point, which is further analyzed in Ref. [104]. By contrast, the states within the $|Q; 0_+\rangle$ family are always degenerate with those of the $|Q; 0_-\rangle$ family, as we have already mentioned in Sec. 3.4. Correspondingly, for every value of Q, each of the dotted lines in the figures simultaneouly represents not one, but two states.

The effective 2×2 Hamiltonian for the chiral doublets has been discussed in Refs. [100, 105, 106] and will be presented in Sec. 3.9. The energy of the low-lying $|Q; 0_\pm\rangle$ doublets follows a quadratic dispersion both at small and large momenta, with a crossover taking place around the minimum of the electron-hole continuum, as the inversion of the bands after that has less of an effect. Also, the energies of the higher excited states closely follow the electron-hole continuum, because the excitons are more delocalized in real space and their corresponding wave functions approach those of a separate electron-hole pair. We have checked that the contribution of nonzero angular momenta is important and must be included, as it drives the excitonic dispersion into a clear indirect regime. By contrast, solving the Bethe-Salpeter equation with an ansatz of pure s-wave excitons gives a direct dispersion relation.

Another striking feature is the splitting between the $+m$ and $-m$ states at $Q = 0$ in the sector $s = t$. This is clearly seen in the figures, where for $m \neq 0$ there are always two distinct energy levels with equal $|m|$. We find that the lower one always has $m < 0$ and the higher one has $m > 0$. Note that this is only the case for the states pertaining to the families $|Q; 2_\pm\rangle$. It does not happen for $|Q; 0_\pm\rangle$, for which the zero-momentum states are fourfold degenerate when $|m| > 0$. Since rotational symmetry alone would imply the degeneracy of the $+m$ and $-m$ states at zero momentum, there must exist an additional mechanism producing the observed splitting between these states. This is analyzed in Ref. [102], where it is found that the presence of Berry curvature induces precisely such an energy difference between the $+m$ and $-m$ states. The total Berry curvature is the sum of the electron and hole contributions, $\Omega_{s,t} = \Omega_{e,s} + \Omega_{h,t}$. Note that all functions are evaluated at the momentum k in the case of interest $Q = 0$. In App. 2.A we have determined that $\Omega_{c,s} = -\Omega_{v,s}$ and $\Omega_{c,s} = -\Omega_{c,-s}$. While for fixed spin-orbit parity the Berry curvature of the electron is the same as that of the conduction band, the Berry curvature of the hole is *minus* that of the valence band. This can be seen from our discussion in Sec. 3.1, where we established that the wave function for holes is the complex conjugate of that of valence electrons, and by observing that the Berry curvature is the *imaginary* part of an inner
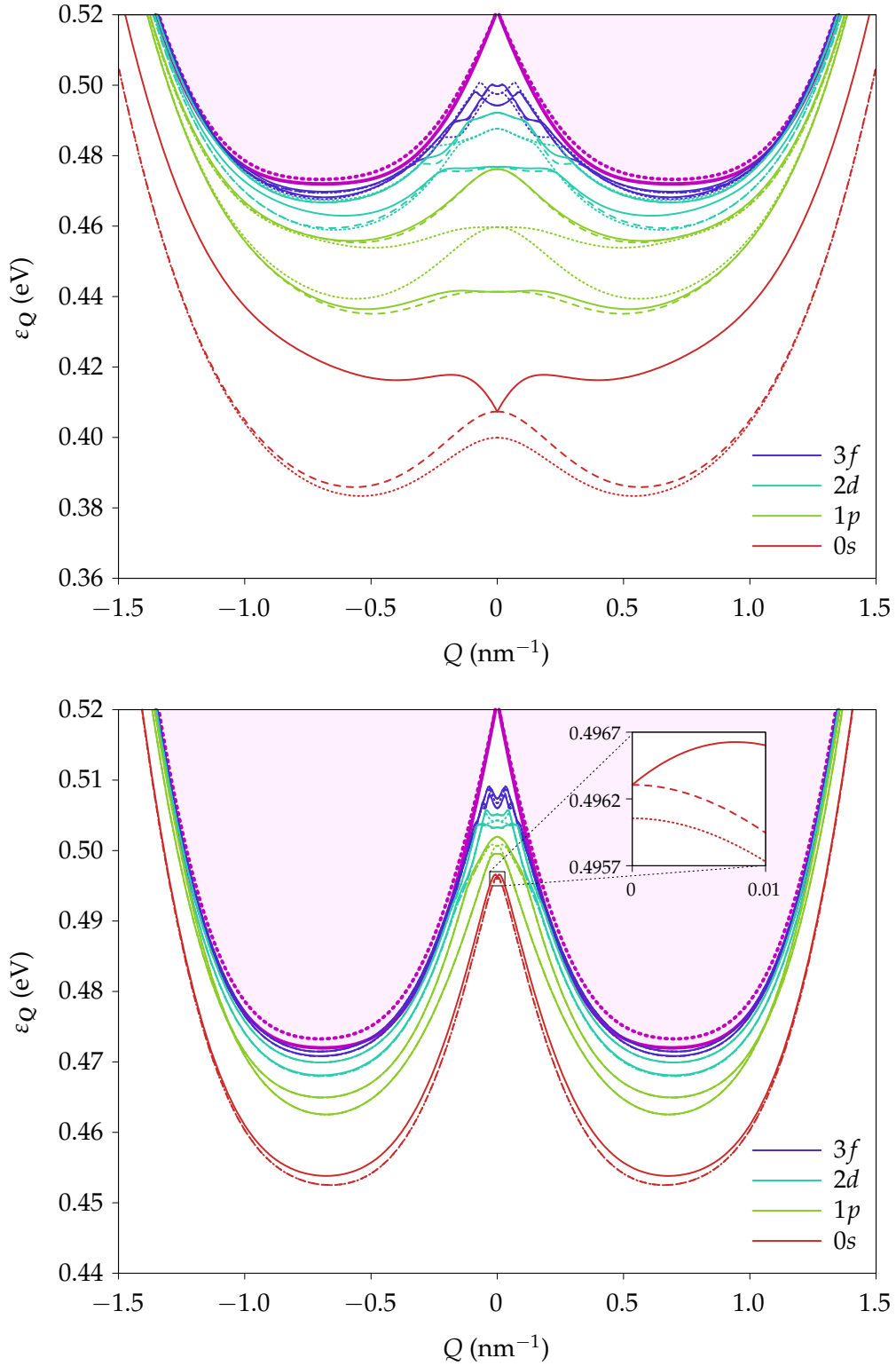
Figure 3.2. Excitonic dispersion relation as a function of the total exciton momentum Q for a few low-lying states as obtained for the bare Coulomb potential with $\varepsilon_s = 6$ (top) and for the Struve-Neumann potential with $\varepsilon_s = 6$ and $\varepsilon_d = 28$ (bottom). Shown are the exciton states $|Q; 0_{\pm}\rangle$ (dotted lines), $|Q; 2_{+}\rangle$ (solid lines), and $|Q; 2_{-}\rangle$ (dashed lines) for different values of m. In both cases the $1s$ state (not shown) has higher energy than the $1p$, $2d$ and $3f$ states. In the subspace with $s = t$, the zero-momentum states with $m > 0$ have larger energy than those with $m < 0$. This splitting arises due to Berry-curvature effects. The magenta region at the top represents the electron-hole continuum, delimited by a solid purple line resulting from minimizing the energy gap $\Delta_Q(k)$. The purple dotted line represents the ground state in the absence of interactions as calculated numerically by including angular momenta $|m| \leq 3$, and approaches the analytical electron-hole continuum when including higher values of m.

product. It follows that the total Berry curvature is $\Omega_{s,t} = \Omega_{c,s} + \Omega_{c,t}$, which is only nonzero when $s = t$. This explains the zero-momentum features of the different exciton states.

The precise manner in which the Berry curvature causes the splitting can be intuitively understood by using the results of Refs. [107–109]. It is shown that, in the presence of an electric field, the Berry curvature can cause an anomalous transverse velocity of the particles in the nth band,

$$\boldsymbol{v}_n(\boldsymbol{k}) = \boldsymbol{\nabla}_{\boldsymbol{k}} \epsilon_n(\boldsymbol{k}) - q\boldsymbol{E} \times \boldsymbol{\Omega}_n(\boldsymbol{k}). \tag{3.27}$$

Here, q is the charge of the particle, and we recall that $\boldsymbol{\Omega}_n(\boldsymbol{k}) = \boldsymbol{\nabla}_{\boldsymbol{k}} \times \boldsymbol{\mathcal{A}}_n(\boldsymbol{k})$. The first term is the usual expression for the group velocity of a particle in the nth band, whereas the second one is the anomalous contribution. Consider now an exciton with zero total momentum and nonzero angular momentum m. There is an electric field pointing from the positively charged hole to the negatively charged electron. Since our problem is two-dimensional, the Berry-curvature vector lies along the z-direction. When the Berry curvatures of the electron and the hole are the same, the anomalous velocity points in the clockwise (counterclockwise) direction for both particles if the Berry curvature points towards the positive (negative) z-direction. Therefore the states with $+m$ and $-m$ will have different orbital velocities, resulting in an energy splitting between the two. A sketch of the situation is shown in Fig 3.3. By contrast, when the Berry curvatures of the electron and hole are opposite of each other, this effect is canceled.
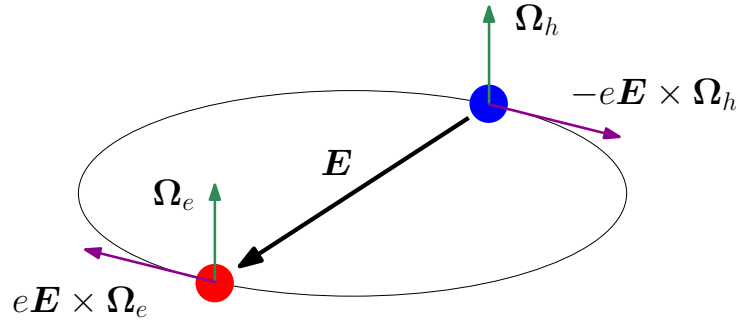

Figure 3.3. Electron (red) and hole (blue) orbiting around each other in a bound state. The orbital velocity is clockwise or counterclockwise depending on the sign of the angular momentum. When the Berry curvatures point in the same direction, the anomalous velocities of both particles also point in the same direction along the orbit. This produces an energy difference between states with opposite angular momenta, as for one of them the orbital speed increases whereas for the other one it is reduced. Figure adapted from Ref. [102].

3.5.3 Spectrum without band inversion

Interestingly, the behavior of the exciton dispersions strongly depends on whether or not the material is in the topologically nontrivial inverted regime. We have repeated the Coulomb-potential calculation in the trivial state by changing the sign of B_2. As shown in Fig. 3.4, we find dispersions which are quadratic in the total exciton momentum, just like we would expect in normal semiconductors. The only exception is a slight linear dependence for the upper mode of the $s = t$ subspace, which reflects the aforementioned effect of the exchange interaction around zero momentum. Furthermore, the energy spectrum for $\boldsymbol{Q} = \boldsymbol{0}$ in each of the four distinct spin-orbit subspaces is analogous to that of the two-dimensional hydrogen atom. The most prominent feature of this spectrum is the accidental $(2n + 1)$fold degeneracy of the nth level, which does not explicitly include the azimuthal quantum number m [101]. Since we have four different combinations of s and t, in our case we obtain a $4(2n + 1)$fold degeneracy for each level (in practice, we numerically find splittings of the order of 0.1 meV). The binding energies

Δ of the zero-momentum states satisfy the rule $\frac{\Delta_n}{\Delta_{n+1}} = \frac{(n+3/2)^2}{(n+1/2)^2}$, which agrees with the spectrum of the 2D hydrogen atom. For nonzero Q, the bands are still found to be almost degenerate, especially for $n \neq 0$. This means that the effect of the exchange interaction is extremely weak in the trivial regime and highlights its importance in a topological phase.
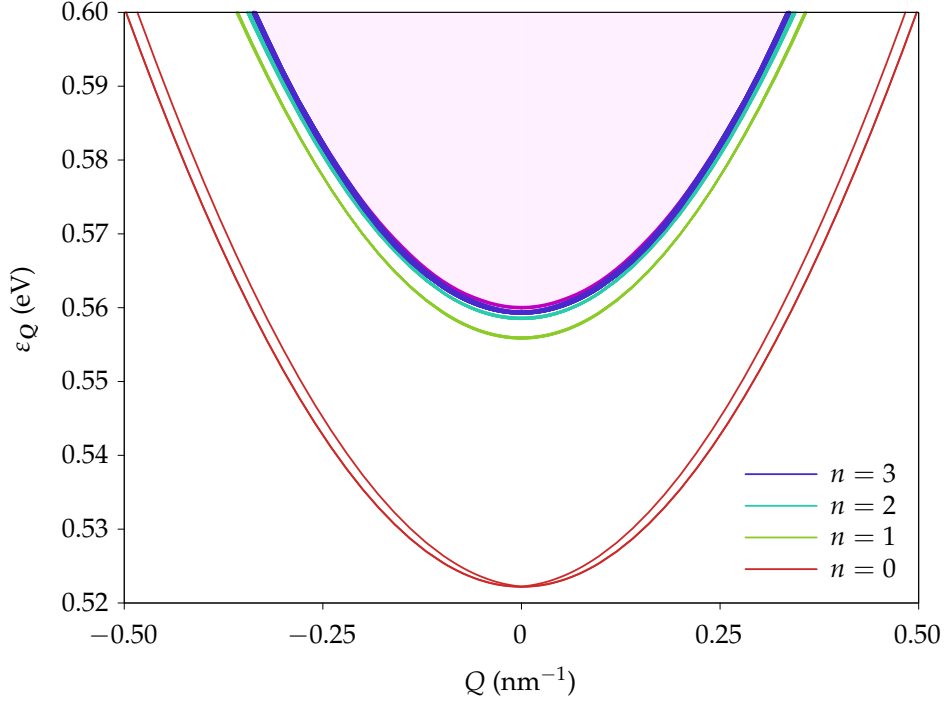

Figure 3.4. Exciton dispersion relations in the trivial insulator phase as obtained for the bare Coulomb potential with $\varepsilon_s = 6$. The states $|0; 2_\pm; n, m\rangle$ and $|0; 0_\pm; n, m\rangle$ with equal n are found to be degenerate at the origin to an accuracy of $\sim 0.1\,\text{meV}$, and there are $4(2n + 1)$ states in each level. The binding energies are proportional to $\left(n + \frac{1}{2}\right)^{-2}$, which precisely matches the two-dimensional hydrogen problem. These results imply that exchange and Berry-curvature effects are weak in comparison to the situation in a topological phase. Once again, the magenta region represents the electron-hole continuum.

3.5.4 Wave functions and probability densities

Solving the BSE also provides us with the exciton momentum-space wave functions for each value of Q. We show the $Q = 0$ Fourier coefficients of the eigenstates $|0; +, +; 0, m\rangle$ in Fig. 3.5, where we have also indicated the binding energies of the corresponding excitons. In particular, we highlight the significantly different behavior of the ground-state wave function compared to that obtained from the hydrogen problem, which is proportional to $(1 + a_0^2 k^2)^{-3/2}$, with a_0 the excitonic Bohr radius.

Furthermore, the momentum-space wave functions can be Fourier-transformed to obtain their real-space counterparts and the corresponding probability densities. Figs. 3.6 and 3.7 show examples of the latter for the subspaces with $s \neq t$ and $s = t$, respectively, for zero and nonzero total exciton momentum. We note that the states $|0; 0_\pm\rangle$ do not appear rotationally invariant, but show hydrogen-like orbitals instead. Since at $Q = 0$ the states with $+m$ and $-m$ within the $s \neq t$ subspace are degenerate, it is possible to create linear combinations such that the momentum-space wave functions are purely real, e.g.,

$$|0; +, -; n, x\rangle \qquad = \frac{1}{\sqrt{2}}\Big(|0; +, -; n, -1\rangle + |0; +, -; n, +1\rangle\Big), \qquad (3.28a)$$
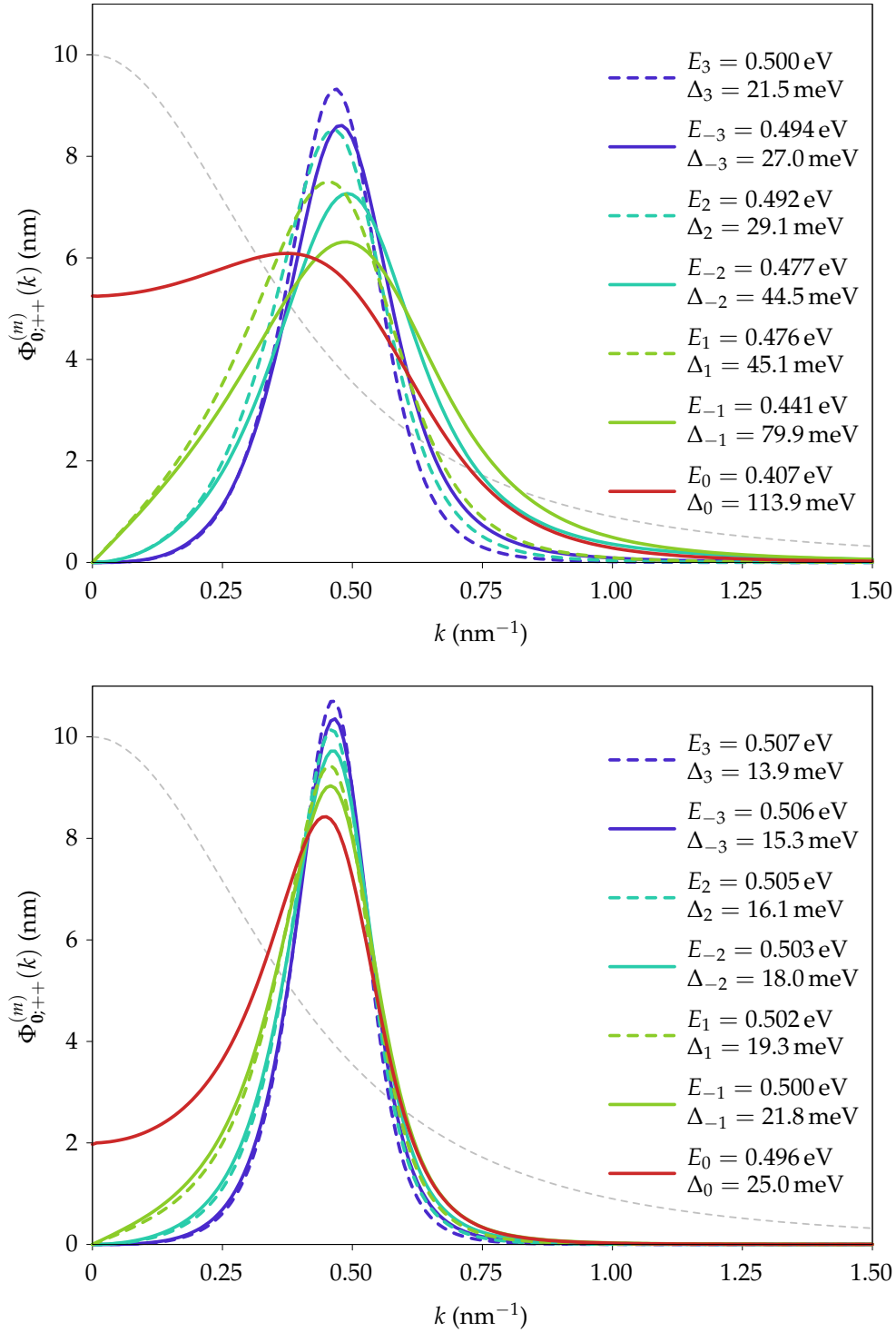
Figure 3.5. Fourier coefficients of the relative momentum-space exciton wave functions of the zero-momentum states $|0; +, +; 0, m\rangle$ and $|0; -, -; 0, -m\rangle$ for several values of m as ordered in Fig. 3.2. Again, the top figure corresponds to the Coulomb potential and the bottom one to the Struve-Neumann potential. The wave functions themselves are obtained by multiplying the magnitude by the corresponding phase $e^{im\phi_k}$. The wave functions have a maximum around the momentum at which the energy gap $\Delta_Q(k)$ presents a minimum. In particular, note that in the s-wave case (red line) the wave function significantly differs from that of the two-dimensional hydrogen atom, which is $(8\pi a_0^2)^{1/2}(1 + a_0^2 k^2)^{-3/2}$ (grey dashed line, with a value of $a_0 = 10/\sqrt{8\pi}$ suitable for comparison). The values E_m shown in the figure correspond to the excitonic eigenenergies of the displayed states and are measured with respect to the chemical potential. The quantities Δ_m are the binding energies of each state, that is, the difference between the electron-hole continuum and E_m.

53

$$|\mathbf{0};+,-;n,y\rangle \quad = \frac{i}{\sqrt{2}}\Big(|\mathbf{0};+,-;n,-1\rangle - |\mathbf{0};+,-;n,+1\rangle\Big), \quad (3.28b)$$

$$|\mathbf{0};+,-;n,x^2-y^2\rangle = \frac{1}{\sqrt{2}}\Big(|\mathbf{0};+,-;n,-2\rangle + |\mathbf{0};+,-;n,+2\rangle\Big), \quad (3.28c)$$

$$|\mathbf{0};+,-;n,2xy\rangle \quad = \frac{i}{\sqrt{2}}\Big(|\mathbf{0};+,-;n,-2\rangle - |\mathbf{0};+,-;n,+2\rangle\Big), \cdots. \quad (3.28d)$$

The different orbitals have been labeled analogously to the hydrogenic case. The same can be done with the states $|\mathbf{0};-,+;n,m\rangle$. Then we can again form doublets $|\mathbf{0};0_\pm\rangle$, which this time do not have a well-defined value of m, but do correspond to a well-defined orbital. This is not possible for the sector with $s = t$, where the $+m$ and $-m$ states are split by the Berry curvature, and consequently their probability densities are rotationally invariant.

Quite generally, we see that the effect of a nonzero exciton momentum is to deform the wave functions due to the breaking of rotational symmetry. Since the longitudinal and transversal directions are not equivalent, the probability densities show lobes in one of these directions.

3.6 Exciton diameters and validity of the first-band approximation

We have also computed the mean excitonic diameter for the low-lying states at zero total momentum, which is a straightforward task once the real-space probability density is known. The results are shown in Table 3.2 for both the Coulomb and Struve-Neumann potentials. We recall that the Coulomb potential is only expected to be accurate in the regime $r \gg (\varepsilon_d/2\varepsilon_s)\ell$, as explained in Sec. 3.3.1. Hence, in this case we must compare the obtained diameters with the nanosheet thickness. Since we chose $\ell = 6\,\text{nm}$, we immediately see that the Coulomb potential cannot reliably model the electron-hole interaction in this system. Indeed, the diameter of the lowest-lying excitons turns out to be *smaller* than the nanosheet thickness, and the limit $r \gg (\varepsilon_d/2\varepsilon_s)\ell$ is not reached.

m		0	-1	$+1$	-2	$+2$	-3	$+3$
$\sqrt{\langle r^2\rangle}$ (nm)	V^{C}	2.74	4.21	5.32	6.33	7.61	9.04	9.76
	V^{SN}	5.64	7.02	7.51	8.72	9.13	10.57	10.89

Table 3.1. Mean exciton diameters of the zero-momentum states $|\mathbf{0};+,+;0,m\rangle$ and $|\mathbf{0};-,-;0,-m\rangle$ as obtained with the Coulomb and Struve-Neumann potentials with $\varepsilon_s = 6$ and $\varepsilon_d = 28$. In the case of the Coulomb potential the radii must be compared to the film thickness, which is $\ell = 6\,\text{nm}$.

Even though this argument suffices to discard the validity of the Coulomb potential in the treatment of this problem, the same cannot be said about the Struve-Neumann interaction. The latter is in principle only valid for $r \gg \ell$, which only seems to be the case for the higher-lying states. However, as we discussed in Sec. 3.3.1, the SN potential is very close to the quantum-confined Rytova-Keldysh potential, which in any case is expected to be the most accurate choice for the system under consideration. Therefore, a diameter smaller or of the order of the nanosheet thickness is of no concern in the case of the SN potential.

Consequently, the validity of the Struve-Neumann potential is solely determined by the procedure anticipated in Sec. 2.4.1. We must compare the binding energies of the excitons with the splitting of the first two bulk subbands. The splittings of the conduction and valence subbands for $\ell = 6\,\text{nm}$ are readily obtained from Table 2.2 and are about 61 meV and 24 meV, respectively. The binding energies are given in Fig. 3.5, but we repeat them in the table below.

We see that, except for the ground state, all states have a binding energy that is smaller than the relevant subband splittings. In the case of the ground state, the binding energy is
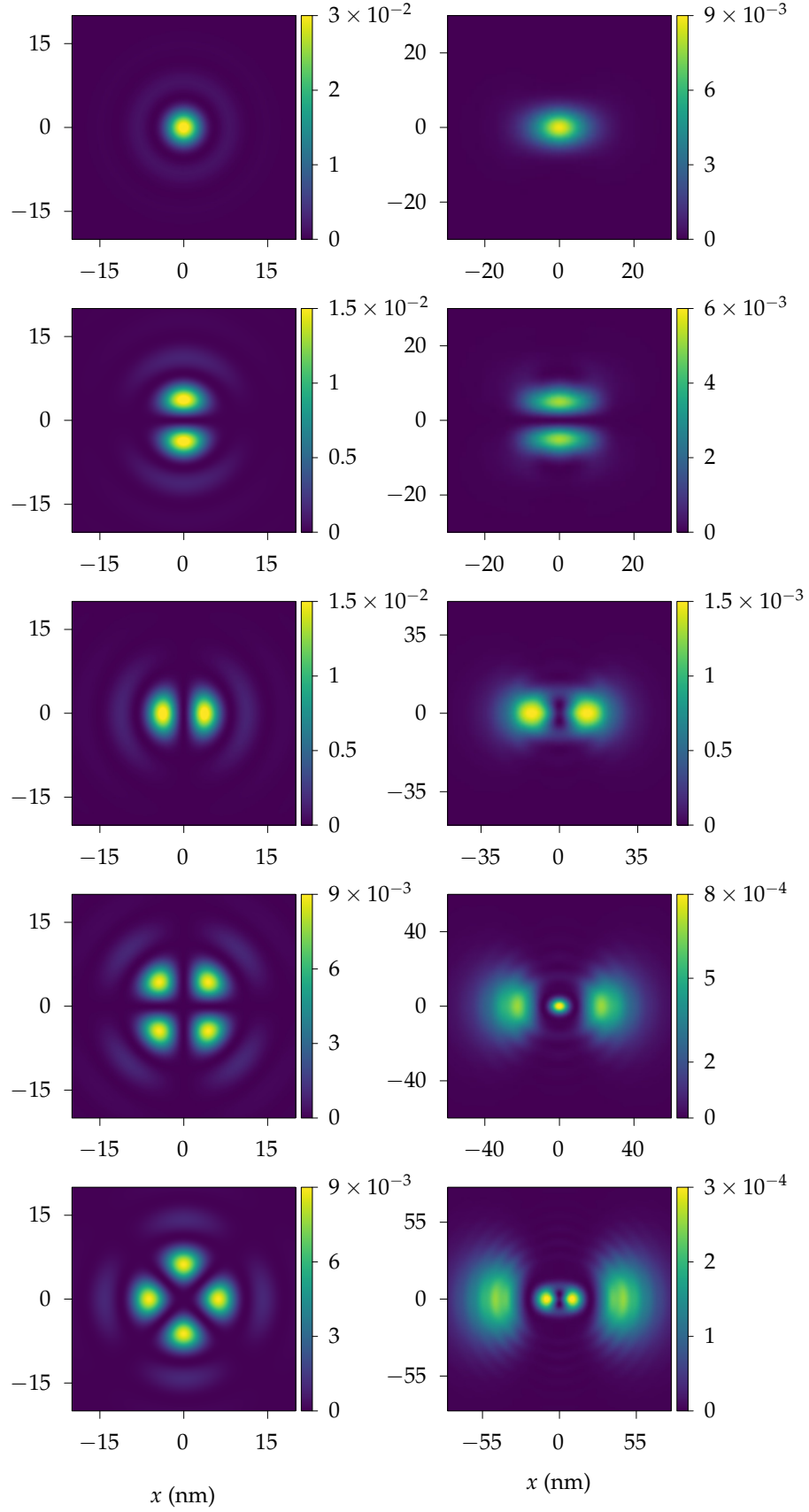
Figure 3.6. Real-space probability densities of linear combinations of the states $|0; 0_\pm; 0, m\rangle$ (left) and of the states $|Q; 0_\pm; 0, m\rangle$ (right), with Q along the x-direction and $Q = 0.7\,\text{nm}^{-1}$. The zero-momentum states represented here do not have a well-defined angular momentum, but rather show the orbitals introduced in Eqs. (3.28), namely (from top to bottom): 1, y, x, $2xy$, $x^2 - y^2$.
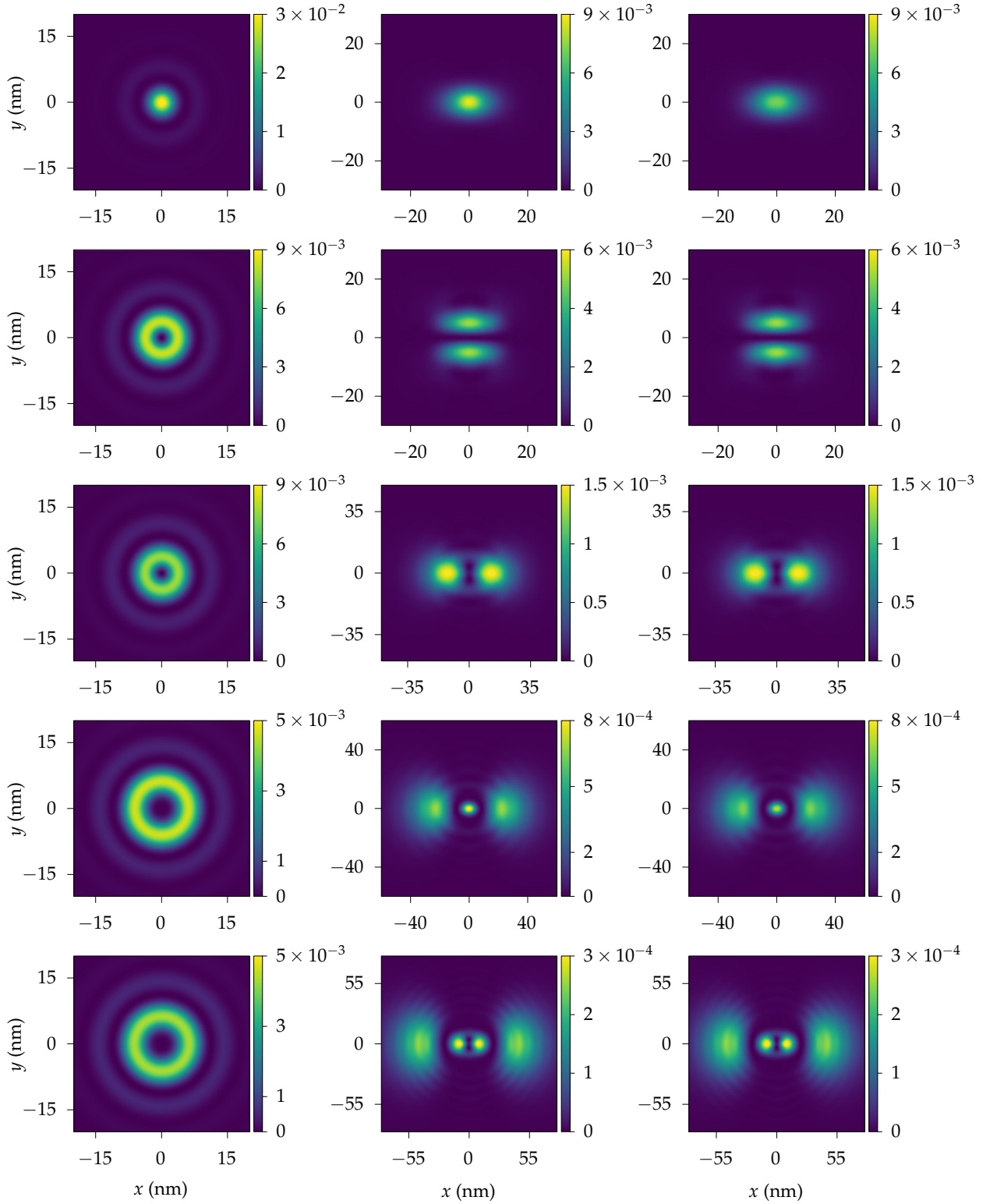
55

Figure 3.7. Real-space probability densities of the states $|0; 2_{\pm}; 0, m\rangle$ (left), $|Q; 2_{-}; 0, m\rangle$ (center) and $|Q; 2_{+}; 0, m\rangle$ (right), with Q along the x-direction and $Q = 0.7 \, \mathrm{nm}^{-1}$. Each row corresponds to a different m as ordered in Fig. 3.2a, namely (from top to bottom): $m = 0, -1, +1, -2, +2$.

m	0	-1	$+1$	-2	$+2$	-3	$+3$
Δ (meV)	25.0	21.8	19.3	18.0	16.1	15.3	13.9

Table 3.2. Binding energies of the zero-momentum states $|0; +, +; 0, m\rangle$ and $|0; -, -; 0, -m\rangle$ as obtained with the Struve-Neumann potential with $\varepsilon_s = 6$ and $\varepsilon_d = 28$.

only around 1 meV larger than the valence-band splitting, so we do not expect that including the second subband will lead to great modifications. We conclude that the Struve-Neumann potential is indeed apt for our study of excitons in Bi$_2$Se$_3$. On a final note, even though we have only presented the analysis for the subspace with $s = t$, for $s \neq t$ the results are extremely similar and all conclusions hold in this case as well.

3.7 Selection rules and exciton brightness

Let us now study the selection rules for the generation of excitons in our model. We recall that interband transitions between Bi and Se orbitals are dipole-allowed, as discussed in Sec. 1.5.3. If Bi$_2$Se$_3$ were a normal semiconductor, we would expect s-wave excitons to be bright and all others to be dark. We will see that this is not the case, and that some s-wave excitons are dark while some d-wave excitons turn out to be bright.

It will be convenient to write the original single-particle Hamiltonian (2.26) in the following form:

$$H(\mathbf{k}) = \epsilon_0(\mathbf{k})\,\mathbb{I}_4 + \mathcal{M}(\mathbf{k})\,\mathbb{I}_2 \otimes \tau_z + A_2(k_x\sigma_x + k_y\sigma_y) \otimes \tau_x, \tag{3.29}$$

where we recall that $\mathcal{M}(\mathbf{k}) = M - B_2(k_x^2 + k_y^2)$. We will focus on excitons with total momentum $\mathbf{Q} = \mathbf{0}$, as these are the most relevant ones from an experimental point of view. We have seen that in this case their angular momentum m is a good quantum number. Furthermore, the electron and hole can be taken to have well-defined spin-orbit parities s and t, respectively.

Consider a beam of light incident on the nanosheet located on the xy-plane. For simplicity we take the photons as propagating in the positive z-direction. Thus, the polarization is parallel to the nanosheet, and is described by a unit Jones vector $\hat{\mathbf{e}}$. The oscillator strength of an exciton $|0; s, t; m\rangle$ reads [103, 110]

$$f_{s,t}^{m,\hat{e}} = \frac{2}{\varepsilon_{s,t}^m} \left| \int \mathrm{d}^2k\, \Phi_{0;s,t}^{(m)} \mathrm{e}^{\mathrm{i}m\phi_k}\, \hat{\mathbf{e}} \cdot \langle \chi_k^{c,s} | \mathbf{v}(\mathbf{k}) | \chi_k^{v,t} \rangle \right|^2. \tag{3.30}$$

Note that we effectively omit the principal quantum number n, as it does not play any role in the discussion of this section. In this expression, $\varepsilon_{s,t}^m$ is the energy of the zero-momentum exciton and $\mathbf{v}(\mathbf{k})$ is the velocity operator.

In general, there are two contributions to the velocity matrix element contained in the formula for the oscillator strength. The first one is interatomic and comes from electron hopping between different lattice sites as described by the underlying tight-binding model. The second one is intraatomic and arises from dipole transitions between orbitals of the same atom [111]. As explained in Sec. 1.4.2, the relevant orbitals of Bi and Se are p orbitals. Optical transitions among them are dipole-forbidden due to the fact that they have the same parity. Therefore, we only expect the interatomic contribution to play a role in Bi$_2$Se$_3$. In this case, it is valid to approximate the velocity operator by

$$\mathbf{v}(\mathbf{k}) = \mathbf{\nabla}_k H(\mathbf{k}). \tag{3.31}$$

We will focus on left- and right-circularly polarized light, which is described by the Jones vectors

$$\hat{e}_\pm = \frac{1}{\sqrt{2}}(\hat{x} \pm i\hat{y}). \tag{3.32}$$

More precisely, $\hat{e}_+$ corresponds to left-circular polarization and $\hat{e}_-$ corresponds to right-circular polarization. Then one finds

$$\hat{e}_\pm \cdot v(k) = \frac{1}{\sqrt{2}} A_2 \sigma_\pm \otimes \tau_x - \sqrt{2} B_2 k_\pm \mathbb{I}_2 \otimes \tau_z, \tag{3.33}$$

where $\sigma_\pm \equiv \sigma_x \pm i\sigma_y$. Note that we have dropped the term proportional to $\partial_{k_i}\epsilon_0(k)$, as it will always drop out of the intraband transitions in Eq. (3.30). One thing to observe is that the operators $\sigma_\pm \otimes \tau_x$ and $\mathbb{I}_2 \otimes \tau_z$ are all spin-orbit-parity-even, that is, they do not couple electrons with different spin-orbit parities. It immediately follows that, at least at the dipole level, all excitons with $s \neq t$ are dark under circularly polarized light, irrespective of their angular momentum. By contrast, the velocity-matrix elements between valence and conduction states with $s = t$ read

$$\langle \chi_k^{c,+} | \hat{e}_+ \cdot v(k) | \chi_k^{v,+} \rangle = \frac{\sqrt{2}}{1 + k^2 f(k)^2} \left[2B_2 k^2 f(k) + A_2 \right], \tag{3.34a}$$

$$\langle \chi_k^{c,+} | \hat{e}_- \cdot v(k) | \chi_k^{v,+} \rangle = \frac{\sqrt{2} k^2 f(k)}{1 + k^2 f(k)^2} \left[2B_2 - A_2 f(k) \right] e^{-2i\phi_k}, \tag{3.34b}$$

$$\langle \chi_k^{c,-} | \hat{e}_+ \cdot v(k) | \chi_k^{v,-} \rangle = \frac{\sqrt{2} k^2 f(k)}{1 + k^2 f(k)^2} \left[2B_2 - A_2 f(k) \right] e^{2i\phi_k}, \tag{3.34c}$$

$$\langle \chi_k^{c,-} | \hat{e}_- \cdot v(k) | \chi_k^{v,-} \rangle = \frac{\sqrt{2}}{1 + k^2 f(k)^2} \left[2B_2 k^2 f(k) + A_2 \right], \tag{3.34d}$$

where $f(k)$ was defined in Eq. (2.29) of the previous chapter. Except for the phase factors $e^{im\phi_k}$ and $e^{\pm 2i\phi_k}$, all the quantities that appear in Eqs. (3.30) and (3.34) are real. Therefore, we can immediately read off which excitons are dark and which ones are bright by checking whether the oscillator strengths are zero or nonzero, respectively. For left-circularly polarized light $\hat{e}_+$ we find that the excitons $|0; +, +; 0\rangle$ and $|0; -, -; -2\rangle$ are bright, whereas the rest are dark. On the other hand, for right-circularly polarized light $\hat{e}_-$ the only optically active excitons are $|0; -, -; 0\rangle$ and $|0; +, +; +2\rangle$. Note that these results combined are in accord with time-reversal symmetry.

As advertised, our findings indicate that optical transitions in Bi_2Se_3 greatly differ from the situation in ordinary semiconductors, where typically s-wave excitons are bright and the rest are dark. Here we have found both bright and dark s-wave excitons, with their response to circular light being subject to the spin-orbit parity of the constituent electrons and holes. Furthermore, some d-wave excitons are also optically active, with the polarization direction and the spin-orbit labels together determining the sign of their azimuthal angular momentum.

3.8 Topology of the exciton basis

In this section we will prove that the exciton basis $|Q; s, t\rangle$ inherits the topology of the underlying band structure that was determined in Sec. 2.5.2. For a state with well-defined s and t, a Chern number can be defined as in the single-particle case. Therefore, in this section we will first ignore the coupling induced by the exchange interaction and compute the Chern number of the exciton states $|Q; s, t\rangle$.

The first step is to obtain the Berry connection, $\mathcal{A}_{s,t}(\boldsymbol{Q}) = -\mathrm{i}\langle \boldsymbol{Q}; s, t | \boldsymbol{\nabla}_{\boldsymbol{Q}} | \boldsymbol{Q}; s, t\rangle$. We find

$$\mathcal{A}_{s,t} = \mathcal{A}_{s,t}^{(e)} + \mathcal{A}_{s,t}^{(h)} + \mathcal{A}_{s,t}^{(\Phi)}, \tag{3.35}$$

where

$$\mathcal{A}_{s,t}^{(e)}(\boldsymbol{Q}) = -\frac{\mathrm{i}}{V} \sum_{\boldsymbol{k}} |\Phi_{\boldsymbol{Q};s,t}(\boldsymbol{k})|^2 \langle \chi_{\boldsymbol{Q}/2+\boldsymbol{k}}^{e,s} | \boldsymbol{\nabla}_{\boldsymbol{Q}} | \chi_{\boldsymbol{Q}/2+\boldsymbol{k}}^{e,s} \rangle, \tag{3.36a}$$

$$\mathcal{A}_{s,t}^{(h)}(\boldsymbol{Q}) = -\frac{\mathrm{i}}{V} \sum_{\boldsymbol{k}} |\Phi_{\boldsymbol{Q};s,t}(\boldsymbol{k})|^2 \langle \chi_{\boldsymbol{Q}/2-\boldsymbol{k}}^{h,t} | \boldsymbol{\nabla}_{\boldsymbol{Q}} | \chi_{\boldsymbol{Q}/2-\boldsymbol{k}}^{h,t} \rangle, \tag{3.36b}$$

$$\mathcal{A}_{s,t}^{(\Phi)}(\boldsymbol{Q}) = -\frac{\mathrm{i}}{V} \sum_{\boldsymbol{k}} \Phi_{\boldsymbol{Q};s,t}^*(\boldsymbol{k}) \boldsymbol{\nabla}_{\boldsymbol{Q}} \Phi_{\boldsymbol{Q};s,t}(\boldsymbol{k}). \tag{3.36c}$$

The first two expressions are simply the $\boldsymbol{k}$-space-averaged Berry curvatures of the electron and the hole, respectively. The term $\mathcal{A}^{(\Phi)}$ represents a contribution arising intrinsically from the relative electron-hole wave function. We will see that this term does not contribute to the Chern number, and hence does not play a significant role in the topology of the problem.

For the Berry curvature $\boldsymbol{\Omega}(\boldsymbol{Q}) = \boldsymbol{\nabla}_{\boldsymbol{Q}} \times \mathcal{A}(\boldsymbol{Q})$ one finds

$$\boldsymbol{\Omega}_{s,t} = \boldsymbol{\Omega}_{s,t}^{(e)} + \boldsymbol{\Omega}_{s,t}^{(h)} + \boldsymbol{\Omega}_{s,t}^{(\Phi)}, \tag{3.37}$$

with

$$\boldsymbol{\Omega}_{s,t}^{(e)}(\boldsymbol{Q}) = -\frac{\mathrm{i}}{V} \sum_{\boldsymbol{k}} \boldsymbol{\nabla}_{\boldsymbol{Q}} \times \left(|\Phi_{\boldsymbol{Q};s,t}(\boldsymbol{k})|^2 \langle \chi_{\boldsymbol{Q}/2+\boldsymbol{k}}^{e,s} | \boldsymbol{\nabla}_{\boldsymbol{Q}} | \chi_{\boldsymbol{Q}/2+\boldsymbol{k}}^{e,s} \rangle \right), \tag{3.38a}$$

$$\boldsymbol{\Omega}_{s,t}^{(h)}(\boldsymbol{Q}) = -\frac{\mathrm{i}}{V} \sum_{\boldsymbol{k}} \boldsymbol{\nabla}_{\boldsymbol{Q}} \times \left(|\Phi_{\boldsymbol{Q};s,t}(\boldsymbol{k})|^2 \langle \chi_{\boldsymbol{Q}/2-\boldsymbol{k}}^{h,t} | \boldsymbol{\nabla}_{\boldsymbol{Q}} | \chi_{\boldsymbol{Q}/2-\boldsymbol{k}}^{h,t} \rangle \right), \tag{3.38b}$$

$$\boldsymbol{\Omega}_{s,t}^{(\Phi)}(\boldsymbol{Q}) = -\frac{\mathrm{i}}{V} \sum_{\boldsymbol{k}} \left(\boldsymbol{\nabla}_{\boldsymbol{Q}} \Phi_{\boldsymbol{Q};s,t}^*(\boldsymbol{k}) \right) \times \left(\boldsymbol{\nabla}_{\boldsymbol{Q}} \Phi_{\boldsymbol{Q};s,t}(\boldsymbol{k}) \right). \tag{3.38c}$$

Let us now calculate the Chern number of the state $|\boldsymbol{Q}; s, t\rangle$. We can use Stokes' theorem to write its expression as

$$\mathcal{C}_{s,t} = \frac{1}{2\pi} \int \mathrm{d}^2\boldsymbol{Q} \cdot \left(\boldsymbol{\nabla}_{\boldsymbol{Q}} \times \mathcal{A}^{(s,t)}(\boldsymbol{Q}) \right) = \frac{1}{2\pi} \oint \mathrm{d}\boldsymbol{Q} \cdot \mathcal{A}^{(s,t)}(\boldsymbol{Q}). \tag{3.39}$$

Note that $\mathrm{d}^2\boldsymbol{Q} \equiv \hat{\boldsymbol{z}}\, \mathrm{d}^2 Q$ is a vector-valued surface element. Similarly, $\mathrm{d}\boldsymbol{Q}$ is a vector-valued line element. The line integral is performed along a circular contour with $Q \to \infty$. We split the total Chern number as $\mathcal{C}_{s,t} = \mathcal{C}_{s,t}^{(e)} + \mathcal{C}_{s,t}^{(h)} + \mathcal{C}_{s,t}^{(\Phi)}$. We first show that $\mathcal{C}_{s,t}^{(\Phi)}$ vanishes. Omitting all irrelevant indices, we have:

$$\begin{aligned}
\mathcal{C}^{(\Phi)} &\propto \int \mathrm{d}^2\boldsymbol{Q} \cdot \left(\boldsymbol{\nabla}_{\boldsymbol{Q}} \Phi_{\boldsymbol{Q}}^* \right) \times \left(\boldsymbol{\nabla}_{\boldsymbol{Q}} \Phi_{\boldsymbol{Q}} \right) \\
&= \int \mathrm{d}^2 R\, \Phi_{\boldsymbol{R}}^* \int \mathrm{d}^2 R'\, \Phi_{\boldsymbol{R}'} \int \mathrm{d}^2\boldsymbol{Q} \cdot \left(\boldsymbol{\nabla}_{\boldsymbol{Q}} \mathrm{e}^{-\mathrm{i}\boldsymbol{R}\cdot\boldsymbol{Q}} \right) \times \left(\boldsymbol{\nabla}_{\boldsymbol{Q}} \mathrm{e}^{\mathrm{i}\boldsymbol{R}'\cdot\boldsymbol{Q}} \right) \\
&= \int \mathrm{d}^2 R\, \Phi_{\boldsymbol{R}}^* \int \mathrm{d}^2 R'\, \Phi_{\boldsymbol{R}'} (\boldsymbol{R} \times \boldsymbol{R}') \cdot \int \mathrm{d}^2\boldsymbol{Q}\, \mathrm{e}^{-\mathrm{i}(\boldsymbol{R}-\boldsymbol{R}')\cdot\boldsymbol{Q}} \\
&= (2\pi)^2 \int \mathrm{d}^2 R\, \Phi_{\boldsymbol{R}}^* \Phi_{\boldsymbol{R}} (\boldsymbol{R} \times \boldsymbol{R})_z \\
&= 0, \tag{3.40}
\end{aligned}$$

where we have defined the Fourier transform $\Phi_Q = \int d^2Q\, \Phi_R\, e^{iR\cdot Q}$ and in the second-to-last step we have used the fact that $\int d^2Q\, e^{-i(R-R')\cdot Q} = (2\pi)^2\delta(R - R')$. This shows that the wave function itself does not contribute to the Chern number.

On the other hand, the electron contribution reads

$$\mathcal{C}_{s,t}^{(e)} = -\frac{1}{2\pi}\frac{i}{V}\sum_{k}\oint dQ \cdot \left(|\Phi_{Q;s,t}(k)|^2\langle \chi_{Q/2+k}^{e,s}|\nabla_Q|\chi_{Q/2+k}^{e,s}\rangle\right). \tag{3.41}$$

This should be equal to the Chern number of a single noninteracting conduction electron. Indeed, we can imagine continuously decreasing the electron-hole interaction by reducing the appropriate coupling constant. The wave function becomes strongly peaked at some value of k, and hence the state $|Q; s, t\rangle$ contains two particles of well-defined momentum. The Chern number is left unchanged under this deformation. More quantitatively, we can argue that this is the case as follows. Since the line integral is performed along a contour at infinity, we can approximate $Q/2 + k \approx Q/2$. Then the only k-dependence in the integrand comes from the wave function, and we can perform the sum $\frac{1}{V}\sum_{k}|\Phi_{Q;s,t}(k)|^2 = 1$. We rename the Chern number to $\mathcal{C}_s^{(e)}$, as it is now manifestly independent of t. Applying Stokes' theorem again we are left with

$$\mathcal{C}_s^{(e)} = -\frac{i}{2\pi}\int d^2Q \cdot \left(\nabla_Q\langle\chi_{Q/2}^{e,s}|\right)\times\left(\nabla_Q|\chi_{Q/2}^{e,s}\rangle\right). \tag{3.42}$$

By performing the change of variable $Q' = Q/2$ we see that $d^2Q = 4d^2Q'$ and $\nabla_Q = \frac{1}{2}\nabla_{Q'}$. Hence we can simply remove the factor of $\frac{1}{2}$ from $|\chi_{Q/2}^{e,s}\rangle$. This is just to say that a rescaling of the momentum cannot change the winding number. What is left is simply

$$\mathcal{C}_s^{(e)} = -\frac{i}{2\pi}\int d^2Q\left(\langle\partial_{Q_x}\chi_Q^{e,s}|\partial_{Q_y}\chi_Q^{e,s}\rangle - \langle\partial_{Q_y}\chi_Q^{e,s}|\partial_{Q_x}\chi_Q^{e,s}\rangle\right). \tag{3.43}$$

But this is precisely the definition of the Chern number of a single noninteracting conduction electron, so that $\mathcal{C}_s^{(e)} = s$ as computed in App. 2.A. On the other hand, for the hole we get

$$\mathcal{C}_t^{(h)} = -\frac{i}{2\pi}\int d^2Q\left(\langle\partial_{Q_x}\chi_Q^{h,t}|\partial_{Q_y}\chi_Q^{h,t}\rangle - \langle\partial_{Q_y}\chi_Q^{h,t}|\partial_{Q_x}\chi_Q^{h,t}\rangle\right). \tag{3.44}$$

In this case we must be careful when using the result from App. 2.A, because the hole wave function is the complex conjugate of the valence-electron wave function. This results in an additional minus sign for the Chern number, in the same way as what was explained at the end of Sec. 3.5.2. Hence we get $\mathcal{C}_t^{(h)} = t$. Therefore, the Chern number of the state $|Q; s, t\rangle$ in the absence of the exchange interaction is

$$\mathcal{C}_{s,t} = s + t. \tag{3.45}$$

As we have said before, switching on the exchange interaction does not couple the subspaces with $s = -t$, and the states $|Q; s, t\rangle$ remain eigenstates of the problem. These states then have a zero Chern number, and are therefore topologically trivial. On the other hand, what happens in the subspace with $s = t$ is more subtle. It is tempting to use Eq. (3.24b) to compute their Chern number, but this is not allowed. The reason is that, despite their coupling, the bands touch at $Q = 0$. Their individual Chern numbers are thus not defined. However, as determined in Sec. 3.4.2, a nontrivial winding of the spin-orbit pseudospin remains. The chiral doublet is therefore expected to be topologically nontrivial.

If we do want to define a Chern number for the chiral doublet, we need to open a small gap at $Q = 0$. This will be done in the next section, where we first derive an effective model for small Q. Despite being an approximation, this model preserves the topology around the Γ point. The simplicity of the model will make it easy to incorporate a gap opening, which in turn will further reveal the topological nontriviality of the $s = t$ subspace.

3.9 Effective model for small Q

We now focus on the subspace with $s = t$, which is the one hosting excitonic states with nontrivial topology. We have seen that, at $Q = 0$, the excitonic spectrum in this subspace consists of degenerate pairs, namely $|0; +, +; n, m\rangle$ and $|0; -, -; n, -m\rangle$. Applying degenerate perturbation theory to these pairs of states (cf. App. 3.E for details), one arrives at the effective Hamiltonian

$$H_{n,m}^{\text{eff}}(Q) = \left(\omega_{n,m} + \frac{Q^2}{2M_{n,m}} \right) \mathbb{I}_2 + \mathcal{J}_{n,m}(Q) \begin{bmatrix} 1 & e^{-2i\phi_Q} \\ e^{2i\phi_Q} & 1 \end{bmatrix}. \tag{3.46}$$

Here, $\omega_{n,m}$ is the energy of the doublet at zero Q and $M_{n,m}$ is an effective mass which can be fit to the numerically obtained band structure. Note that $M_{n,m}$ can be negative, which is the case for the ground state excitons. The coupling $\mathcal{J}_{n,m}(Q)$ reads

$$\mathcal{J}_{n,m}(Q) = \begin{cases} 0 & \text{if } m \text{ is odd}, \\ J_{n,m}|Q| + \dfrac{Q^2}{2M'_{n,m}} + \mathcal{O}(|Q|^3) & \text{if } m \text{ is even}, \end{cases} \tag{3.47}$$

and it is expected that $J_{n,m\neq 0} \ll J_{n,0}$. The above Hamiltonian coincides with the one derived in Refs. [100, 105, 106] after replacing the K and K' valley indices with our spin-orbit parities near the Γ point.

This effective model explains the small-Q features of Fig. 3.2. Firstly, for even m, both subspaces where $s = t$ are coupled to first order in Q and hence the degeneracy at zero momentum is in principle lifted. Omitting the subscripts n and m for clarity, the eigenenergies are found to be

$$\varepsilon_-(Q) = \omega + \frac{Q^2}{2M}, \tag{3.48a}$$

$$\varepsilon_+(Q) = \omega + 2J|Q| + \left(\frac{1}{2M} + \frac{1}{M'} \right) Q^2. \tag{3.48b}$$

One of the modes goes linearly in $|Q|$, while the other one is quadratic around $Q = 0$. This is clearly observed in Fig. 3.2 for the ground-state excitons with zero angular momentum, for which we find $J > 0$ and $M < 0$. In the case of higher even-m states the degeneracy seems to be lifted very weakly, which indicates that the effective couplings of Eq. (3.46) are small. On the other hand, this model predicts that the odd-m states ought to remain degenerate near $Q = 0$, which can also be observed in the figure. This degeneracy is broken by terms cubic in $|Q|$ which have not been included in the perturbative treatment of App. 3.E.

3.9.1 Topological properties

Let us now analyze the topology of the effective Hamiltonian by focusing on the ground-state doublet. Since the bands touch at zero momentum, it is not entirely clear that it can describe a topological *insulator*. To remedy this we add a small gap term $\Delta\sigma_z$ that breaks the time-reversal symmetry. Furthermore, to lowest order we are allowed to only keep terms linear in Q. We also drop all terms that are proportional to the identity, as they cannot influence the topology of the problem. We are left with a reduced toy model whose topological properties should be qualitatively equivalent to those of the original physical system. The Hamiltonian for this toy model reads

$$H(Q) = \begin{bmatrix} \Delta & J|Q|e^{-2i\phi_Q} \\ J|Q|e^{2i\phi_Q} & -\Delta \end{bmatrix}. \tag{3.49}$$

This can be written as $H(Q) = d(Q) \cdot \sigma$, with $d(Q) = (J|Q|\cos 2\phi_Q, J|Q|\sin 2\phi_Q, \Delta)$. This has the form of a "magnetic field" $d(k)$ that couples to the pseudospin σ. Note that the pseudospin operator agrees with the one introduced in Sec. 3.4.2.

We can now analyze the topological properties as a function of the gap parameter Δ. For $\Delta > 0$, we can write nonsingular eigenstates as

$$|v_-\rangle \propto \begin{bmatrix} -J|Q|e^{-2i\phi_Q} \\ \Delta + \sqrt{\Delta^2 + J^2 Q^2} \end{bmatrix} \quad \text{with energy } \varepsilon_-(Q) = -\sqrt{\Delta^2 + J^2 Q^2}, \tag{3.50a}$$

$$|v_+\rangle \propto \begin{bmatrix} \Delta + \sqrt{\Delta^2 + J^2 Q^2} \\ J|Q|e^{2i\phi_Q} \end{bmatrix} \quad \text{with energy } \varepsilon_+(Q) = +\sqrt{\Delta^2 + J^2 Q^2}. \tag{3.50b}$$

For $\Delta < 0$, the nonsingular eigenstates read

$$|v_-\rangle \propto \begin{bmatrix} -\Delta + \sqrt{\Delta^2 + J^2 Q^2} \\ -J|Q|e^{2i\phi_Q} \end{bmatrix} \quad \text{with energy } \varepsilon_-(Q) = -\sqrt{\Delta^2 + J^2 Q^2}, \tag{3.51a}$$

$$|v_+\rangle \propto \begin{bmatrix} J|Q|e^{-2i\phi_Q} \\ -\Delta + \sqrt{\Delta^2 + J^2 Q^2} \end{bmatrix} \quad \text{with energy } \varepsilon_+(Q) = +\sqrt{\Delta^2 + J^2 Q^2}. \tag{3.51b}$$

In both case the Berry curvatures are found to be

$$\Omega_\pm(Q) = \pm \frac{J^2 \Delta}{\left(\Delta^2 + J^2 Q^2\right)^{3/2}}, \tag{3.52}$$

which can in turn be integrated to yield nonzero Chern numbers

$$\mathcal{C}_\pm = \pm \operatorname{sgn} \Delta. \tag{3.53}$$

The limit $\Delta \to 0$ can be taken from either side, but note that in this case there is an ambiguity in the global phase of the eigenstates. For $\Delta \to 0^+$ one finds

$$|v_-\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} -e^{-2i\phi_Q} \\ 1 \end{bmatrix}, \quad |v_+\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ e^{2i\phi_Q} \end{bmatrix}, \tag{3.54}$$

whereas for $\Delta \to 0^-$ one finds

$$|v_-\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -e^{2i\phi_Q} \end{bmatrix}, \quad |v_+\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} e^{-2i\phi_Q} \\ 1 \end{bmatrix}. \tag{3.55}$$

The Berry curvatures reduce to magentic monopoles in momentum space, namely

$$\Omega_\pm(Q) = \pm 2\pi \operatorname{sgn} \Delta \, \delta(Q). \tag{3.56}$$

Here, $\operatorname{sgn} \Delta$ is defined according to whether the limit is being taken from above or from below. When Δ is *strictly* zero, the individual Chern numbers of the bands are technically speaking not well-defined. However, the most sensible choice in this case is to write down eigenstates for which the usual procedure for the calculation of the Chern number yields zero, that is,

$$|v_\pm\rangle = \frac{1}{\sqrt{2}} \begin{bmatrix} e^{-i\phi_Q} \\ \pm e^{i\phi_Q} \end{bmatrix}. \tag{3.57}$$

Note that this is in fact consistent with defining the sign function at zero as $\operatorname{sgn} 0 = 0$, which is a common convention. The way in which we have written these eigenstates agrees with the chiral doublets of Eq. (3.24b).

3.9.2 Exciton edge states

We have seen that opening a small gap in the dispersion of the ground-state chiral doublet gives a nonzero Chern number for the eigenstates. Let us now consider a position-dependent gap parameter Δ defined as

$$\Delta(x) = \begin{cases} \Delta_1 < 0 & \text{for } x < 0, \\ \Delta_2 > 0 & \text{for } x > 0. \end{cases} \tag{3.58}$$

The Chern numbers of the states $|v_-\rangle$ and $|v_+\rangle$ are therefore opposite at both sides of $x = 0$. Due to the bulk-boundary correspondence, we expect topological surface states of excitons at this interface. These exciton egde states should be chiral due to the breaking of time-reversal symmetry, and we expect two such states due to the difference in Chern numbers at both sides of the interface. A similar problem has been studied by Gong et al. in Ref. [112]. In this article, the gap term is taken as a slowly varying magnetic field of the form $\Delta(x) = \Delta_0 \tanh(x/l)$ and two topological exciton states are indeed identified at the boundary.

We have attempted to solve the simpler model with $\Delta(x)$ as in Eq. (3.58). The substitution $Q_x \to -i\partial_x$ and the requirement that the wave function be normalizable and continuous at the interface ultimately yield the following implicit equation for the energy levels:

$$\frac{\left(|Q| + \sqrt{Q^2 + \Delta_1^2 - E^2}\right)^2}{(E + \Delta_1)\sqrt{E^2 - \Delta_1^2}} = \frac{\left(|Q| - \sqrt{Q^2 + \Delta_2^2 - E^2}\right)^2}{(E + \Delta_2)\sqrt{E^2 - \Delta_2^2}}. \tag{3.59}$$

However, it appears that this equation only has solutions when the signs of Δ_1 and Δ_2 are equal instead of opposite. Furthermore, in this case there seems to be only one solution instead of two. Hence, this procedure apparently does not give the exciton states we are looking for. The resolution of this issue is left for future work.

Appendices

3.A Evaluation of Feynman diagrams

Here we explicitly calculate the Feynman diagrams in Sec. 3.3. The conduction and valence wave functions corresponding to the single-particle states read

$$\langle x; \mathbf{i} | q; c, s \rangle = \frac{e^{iq \cdot x}}{\sqrt{V}} \langle \mathbf{i} | \chi_q^{c,s} \rangle, \tag{3.60a}$$

$$\langle x; \mathbf{i} | q; v, t \rangle = \frac{e^{iq \cdot x}}{\sqrt{V}} \langle \mathbf{i} | \chi_q^{v,t} \rangle, \tag{3.60b}$$

where s and t are the spin-orbit parities and the index $\mathbf{i} \equiv (i, \sigma)$ now denotes our combined spin and orbit subspaces. The (conduction) electron and hole wave functions are then related to the above through

$$\langle x; \mathbf{i} | q; e, s \rangle = \langle x; \mathbf{i} | q; c, s \rangle, \tag{3.61a}$$
$$\langle x; \mathbf{i} | q; h, t \rangle = \langle x; \mathbf{i} | -q; v, t \rangle^*, \tag{3.61b}$$

as explained in Sec. 3.1.

We begin with the direct diagram, which we denote by $V^D(Q, Q'; k, k')$:
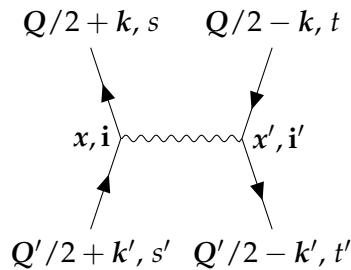
Note that the particles are taken as a electrons and holes, and *not* as valence and conduction electrons. The standard Feynman rules with a spin-orbit-independent interaction V give:

$$
V_{s,t;s',t'}^{D}(Q, Q'; k, k')
$$
$$
= \sum_{\mathbf{i}, \mathbf{i'}} \int d^2x \int d^2x' \, \langle x; \mathbf{i} | Q/2 + k; e, s \rangle^* \langle x'; \mathbf{i'} | Q/2 - k; h, t \rangle^* V(x - x')
$$

$$\times \langle x;\mathbf{i}|Q'/2+k';e,s'\rangle\langle x';\mathbf{i}'|Q'/2-k';h,t'\rangle$$

$$=\sum_{\mathbf{i},\mathbf{i}'}\int \mathrm{d}^2x\int \mathrm{d}^2x'\,\langle x;\mathbf{i}|Q/2+k;c,s\rangle^*\langle x';\mathbf{i}'|-Q/2+k;v,t\rangle\,V(x-x')$$

$$\times \langle x;\mathbf{i}|Q'/2+k';c,s'\rangle\langle x';\mathbf{i}'|-Q'/2+k';v,t'\rangle^*$$

$$=\sum_{\mathbf{i},\mathbf{i}'}\int \mathrm{d}^2x\int \mathrm{d}^2x'\,\frac{\mathrm{e}^{-\mathrm{i}(Q/2+k)\cdot x}}{\sqrt{V}}\frac{\mathrm{e}^{-\mathrm{i}(Q/2-k)\cdot x'}}{\sqrt{V}}\frac{\mathrm{e}^{\mathrm{i}(Q'/2+k')\cdot x}}{\sqrt{V}}\frac{\mathrm{e}^{\mathrm{i}(Q'/2-k')\cdot x'}}{\sqrt{V}}\,V(x-x')$$

$$\times \langle \chi^{c,s}_{Q/2+k}|\mathbf{i}\rangle\langle \mathbf{i}|\chi^{c,s'}_{Q'/2+k'}\rangle\langle \chi^{v,t'}_{-Q'/2+k'}|\mathbf{i}'\rangle\langle \mathbf{i}'|\chi^{v,t}_{-Q/2+k}\rangle$$

$$=\frac{1}{V^2}\,\langle \chi^{c,s}_{Q/2+k}|\chi^{c,s'}_{Q'/2+k'}\rangle\langle \chi^{v,t'}_{-Q'/2+k'}|\chi^{v,t}_{-Q/2+k}\rangle$$

$$\times \int \mathrm{d}^2x\int \mathrm{d}^2x'\,\mathrm{e}^{-\mathrm{i}(Q-Q')\cdot(x+x')/2}\,\mathrm{e}^{-\mathrm{i}(k-k')\cdot(x-x')}\,V(x-x')$$

$$=\frac{1}{V^2}\,\langle \chi^{c,s}_{Q/2+k}|\chi^{c,s'}_{Q'/2+k'}\rangle\langle \chi^{v,t'}_{-Q'/2+k'}|\chi^{v,t}_{-Q/2+k}\rangle\int \mathrm{d}^2R\,\mathrm{e}^{-\mathrm{i}(Q-Q')\cdot R}\int \mathrm{d}^2r\,V(r)\,\mathrm{e}^{-\mathrm{i}(k-k')\cdot r}$$

$$=\frac{1}{V}\,\delta_{Q,Q'}V(k-k')\langle \chi^{c,s}_{Q/2+k}|\chi^{c,s'}_{Q'/2+k'}\rangle\langle \chi^{v,t'}_{-Q'/2+k'}|\chi^{v,t}_{-Q/2+k}\rangle. \tag{3.62}$$

In the third step we have used the fact that $\sum_{\mathbf{i}}|\mathbf{i}\rangle\langle \mathbf{i}|=1$. In the next one we have changed to the relative variables $R=\frac{1}{2}(x+x')$ and $r=x-x'$; note that this transformation has a unit Jacobian. Then we have used the relation $\frac{1}{V}\int \mathrm{d}^2R\,\mathrm{e}^{-\mathrm{i}(Q-Q')\cdot R}=\delta_{Q,Q'}$. As expected, the conservation of the total exciton momentum directly arises from the diagram. Note that the factor of $\frac{1}{V}$ has been omitted in Eq. (3.7a), but reinserted explicitly in (3.16). The expression (3.62) is the same as (3.7a) once we observe that the inner products between two states are orthogonal for any values of the momenta if the spin-orbit parities are different.

Eq. (3.62) suggests the Feynman rules in momentum space. An electron with momentum q and spin-orbit parity s gives $|\chi^{c,s}_q\rangle$ in the initial state and $\langle \chi^{c,s}_q|$ in the final state. A hole with momentum q and spin-orbit parity t gives $\langle \chi^{v,t}_{-q}|$ in the initial state and $|\chi^{v,t}_{-q}\rangle$ in the final state, consistent with the transformation between holes and valence electrons. The interaction comes with a factor of $V(q)/V$, where q is the momentum exchange, and one takes the inner product between the two states at each end of the interaction.

These rules make it simple to guess the expression for the exchange diagram:
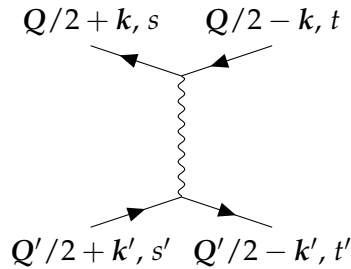
According to what we said, we obtain

$$V^X_{s,t;s',t'}(Q,Q';k,k')=\frac{1}{V}\delta_{Q,Q'}V(Q)\langle \chi^{c,s}_{Q/2+k}|\chi^{v,t}_{-Q/2+k}\rangle\langle \chi^{v,t'}_{-Q'/2+k'}|\chi^{c,s'}_{Q'/2+k'}\rangle. \tag{3.63}$$

A direct calculation similar to the one above indeed yields the exact same result.

3.B Numerical solution of the BSE

After performing a Fourier expansion, the BSE equation reads

$$\sum_{m'}\left[\mathcal{K}^{(m,m')}(\boldsymbol{Q};k)\,\Phi^{(m')}_{\boldsymbol{Q};s,t}(k)+\sum_{s',t'}\int_0^\infty\frac{\mathrm{d}k'}{2\pi}\,k'\mathcal{V}^{(m,m')}_{s,t;s',t'}(\boldsymbol{Q};k,k')\,\Phi^{(m')}_{\boldsymbol{Q};s',t'}(k')\right]=\varepsilon_{\boldsymbol{Q}}\,\Phi^{(m)}_{\boldsymbol{Q};s,t}(k),\quad(3.64)$$

where

$$\mathcal{K}^{(m,m')}(\boldsymbol{Q};k)=\int_0^{2\pi}\frac{\mathrm{d}\phi}{2\pi}\,\mathrm{e}^{-\mathrm{i}(m-m')\phi}\left[\epsilon_c(\boldsymbol{Q}/2+\boldsymbol{k}_\phi)-\epsilon_v(\boldsymbol{Q}/2-\boldsymbol{k}_\phi)\right],\qquad(3.65a)$$

$$\mathcal{V}^{(m,m')}_{s,t;s',t'}(\boldsymbol{Q};k,k')=\int_0^{2\pi}\frac{\mathrm{d}\phi}{2\pi}\int_0^{2\pi}\frac{\mathrm{d}\phi'}{2\pi}\,\mathrm{e}^{-\mathrm{i}m\phi}\mathrm{e}^{\mathrm{i}m'\phi'}\left(V^{\mathrm{D}}-V^{\mathrm{X}}\right)_{s,t;s',t'}(\boldsymbol{Q};\boldsymbol{k}_\phi,\boldsymbol{k}'_{\phi'}),\qquad(3.65b)$$

with $\boldsymbol{k}_\phi=(k\cos\phi,k\sin\phi)$. The unknowns in this equation are the Fourier coefficients $\Phi_{\boldsymbol{Q};s,t}(\boldsymbol{k})$ and the eigenenergies $\varepsilon_{\boldsymbol{Q}}$. For convenience we will work with the dimensionless momentum $u\equiv k\ell$, which requires us to rescale the parameters in Table 2.4 accordingly.

In order to solve the BSE we follow the method described in Ref. [99]. The integral is discretized into a sum over $N+1$ points with momenta $u_n=n\Delta u$, where $n=0,\dots,N$ and $\Delta u\ll 1$ is a small enough momentum step. In other words, we only perform the integral up to some ultraviolet cutoff $U\equiv u_N\gg 1$. The Hamiltonian matrix is firstly divided into blocks of fixed s,t,s' and t'. In practice, we only need to separately consider the Hamiltonians with $s=t$ and $s\neq t$, as they are uncoupled from each other. Furthermore, for $s\neq t$ it suffices to consider the block with $s=+1$, as it is degenerate and uncoupled from the one with $s=-1$. Each such block is further organized into subblocks of fixed m and m', and we have chosen to keep only $|m|\leq 3$ in our numerical calculations. Each subblock is calculated for all values of $u,u'\leq U$, resulting in the matrix elements[5]

$$H^{(m,m')}_{s,t;s',t'}(\boldsymbol{Q};u,u')=\mathcal{K}^{(m,m')}(\boldsymbol{Q};u)\,\delta_{s,t}\delta_{s',t'}\delta_{u,u'}+\frac{\Delta u}{2\pi}u\mathcal{V}^{(m,m')}_{s,t;s',t'}(\boldsymbol{Q};u,u').\qquad(3.66)$$

The Fourier coefficients are accordingly arranged into a single column vector. The Hamiltonian matrix can then be readily diagonalized. This yields the desired eigenenergies, as well as a series of eigenvectors from which the eigenfunctions $\Phi_{\boldsymbol{Q};s,t}(\boldsymbol{k})$ can be recovered by appropriately reconstructing the Fourier sum.

Care must be taken when choosing Δu and U, as these parameters will determine the convergence of this method. However, the problem we are treating actually has an additional degree of freedom besides the discretization and cutoff. The reason is that, in the absence of screening, we need to regularize the interaction potential at zero momentum, where we encounter the Coulomb divergence for any of the potentials of interest. To do this, we introduce an additional parameter $\Delta_V u\leq\Delta u$, which acts as an infrared cutoff such that the interaction for $u<\Delta_V u$ simply saturates to $V(u)\propto(\Delta_V u)^{-1}$.

The continuum limit is achieved when $\Delta_V u\leq\Delta u\to 0$ and $U\to\infty$. Hence, the cutoff U must be chosen large enough for the potential to decrease significantly, which can always be checked *a posteriori* by verifying that the wave functions decay to zero for $u\ll U$. If this condition is met, the energies obtained for a given discretization are found to be independent of the cutoff. By contrast, the energies in general depend slightly on Δu and $\Delta_V u$, but in such a way which allows unambiguous extrapolation to $\Delta u\to 0$.

We have performed calculations for $\boldsymbol{Q}=\boldsymbol{0}$ for different values of Δu and $\Delta_V u$. In Fig. 3.8, we show the numerically obtained ground-state energies of the $s=t$ subspace as a function

---

[5]It is not necessary to calculate every single subblock, as the potentials satisfy some symmetries that relate blocks with different angular momenta. These properties are summarized in the next section.

of the ratio $\Delta_V u / \Delta u$ for fixed Δu, and as a function of Δu for a fixed value of this ratio. These plots have been obtained for the Struve-Neumann potential and with $U = 10$ everywhere, and we have checked that the results are the same for higher values of the cutoff. The first plot reveals that there exists an ideal value of $\Delta_V u / \Delta u$, namely $\Delta_V u / \Delta u \approx 0.2262$, for which the energies are independent of Δu. The second plot shows the energies obtained with this ideal ratio, which indeed lie on a horizontal line. For other values of $\Delta_V u / \Delta u$, the ground-state energies lie on a straight line which can be unambiguously extrapolated to $\Delta u \to 0$, yielding the same energy in all cases. We have verified that the same ideal value of $\Delta_V u / \Delta u$ is found for the subspaces with $s \neq t$.

For our subsequent numerical calculations with the Struve-Neumann potential we have therefore set $U = 10$, $\Delta u = 0.05$ (corresponding to $N = 200$) and $\Delta_V u / \Delta u$ to its ideal value, which allows us to skip the extrapolation to $\Delta u \to 0$.
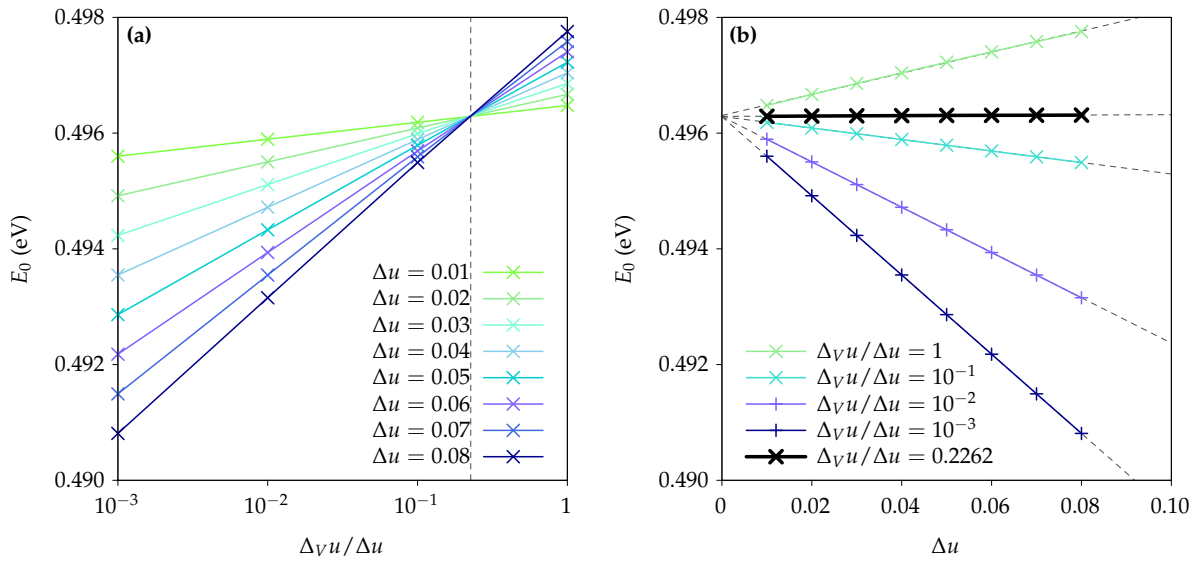

Figure 3.8. (a) Ground-state energy of the $s = t$ subspace as a function of the ratio $\Delta_V u / \Delta u$ for different values of the discretization Δu. Around the ideal value $\Delta_V u / \Delta u \approx 0.2262$ (marked with a vertical dashed line) the obtained energy is independent of the discretization as long as the latter is small enough. **(b)** Ground-state energy as a function of the discretization Δu for different values of the ratio $\Delta_V u / \Delta u$. As $\Delta u \to 0$, all lines tend to the same value independently of $\Delta_V u$. The thicker black line corresponds to a value of $\Delta_V u / \Delta u$ close to the ideal ratio, for which the energies do not depend on Δu, resulting in a horizontal line. In these figures, the Struve-Neumann potential and a cutoff $U = 10$ have been used.

3.C   Symmetries of the potential matrix elements

In order to speed up the numerics we can exploit the symmetries of the potentials to relate certain blocks $H_{s,t;s',t'}^{(m,m')}(\boldsymbol{Q};k,k')$ with different m, m'. First and foremost, we can always use the Hermiticity property

$$\mathcal{V}_{s,t;s',t'}^{(m,m')}(\boldsymbol{Q};k,k') = \mathcal{V}_{s',t';s,t}^{(m',m)}(\boldsymbol{Q};k',k)^* \tag{3.67}$$

when filling the Hamiltonian matrix. However, there exist additional relations between the different subblocks of the potential matrix. We begin by recalling that

$$V_{s,t;s',t'}^{D/X}(\boldsymbol{Q};k,k') = V_{-s,-t;-s',-t'}^{D/X}(\boldsymbol{Q};k,k')^*, \tag{3.68}$$

where $s, t, s', t' \in \{+, -\}$.

Let us start with the subspaces where $s \neq t$. The following properties hold:

$$V^{\mathrm{D}}_{+-;+-}(Q\hat{\boldsymbol{x}}; \boldsymbol{k}_\phi, \boldsymbol{k}'_{\phi'}) = V^{\mathrm{D}}_{+-;+-}(Q\hat{\boldsymbol{x}}; \boldsymbol{k}_{-\phi}, \boldsymbol{k}'_{-\phi'})^*, \tag{3.69a}$$

$$V^{\mathrm{D}}_{+-;+-}(\boldsymbol{Q}; \boldsymbol{k}_\phi, \boldsymbol{k}'_{\phi'}) = V^{\mathrm{D}}_{+-;+-}(\boldsymbol{Q}; \boldsymbol{k}_{\phi+\pi}, \boldsymbol{k}'_{\phi'+\pi})^*. \tag{3.69b}$$

The equality (3.69a) implies that $\mathcal{V}^{(m,m')}_{+-;+-}(Q\hat{\boldsymbol{x}}; k, k')$ is real:

$$\begin{aligned}
\mathcal{V}^{(m,m')}_{+-;+-}(Q\hat{\boldsymbol{x}}; k, k') &= \int_{\phi,\phi'} \mathrm{e}^{-\mathrm{i}m\phi}\mathrm{e}^{\mathrm{i}m'\phi'} V^{\mathrm{D}}_{+-;+-}(Q\hat{\boldsymbol{x}}; \boldsymbol{k}_\phi, \boldsymbol{k}'_{\phi'}) \\
&= \int_{\phi,\phi'} \mathrm{e}^{-\mathrm{i}m\phi}\mathrm{e}^{\mathrm{i}m'\phi'} V^{\mathrm{D}}_{+-;+-}(Q\hat{\boldsymbol{x}}; \boldsymbol{k}_{-\phi}, \boldsymbol{k}'_{-\phi'})^* \\
&= \left[\int_{\phi,\phi'} \mathrm{e}^{\mathrm{i}m\phi}\mathrm{e}^{-\mathrm{i}m'\phi'} V^{\mathrm{D}}_{+-;+-}(Q\hat{\boldsymbol{x}}; \boldsymbol{k}_{-\phi}, \boldsymbol{k}'_{-\phi'}) \right]^* \\
&= \left[\int_{\phi,\phi'} \mathrm{e}^{-\mathrm{i}m\phi}\mathrm{e}^{\mathrm{i}m'\phi'} V^{\mathrm{D}}_{+-;+-}(Q\hat{\boldsymbol{x}}; \boldsymbol{k}_\phi, \boldsymbol{k}'_{\phi'}) \right]^* \\
&= \mathcal{V}^{(m,m')}_{+-;+-}(Q\hat{\boldsymbol{x}}; k, k')^*. \tag{3.70}
\end{aligned}$$

Here and below, the notation $\int_{\phi,\phi'}$ stands for $\int_0^{2\pi} \frac{\mathrm{d}\phi}{2\pi} \int_0^{2\pi} \frac{\mathrm{d}\phi'}{2\pi}$ and $\boldsymbol{k}_\phi \equiv (k\cos\phi, k\sin\phi)$. On the other hand, using (3.69b) we get

$$\begin{aligned}
\mathcal{V}^{(-m,-m')}_{+-;+-}(\boldsymbol{Q}; k, k') &= \int_{\phi,\phi'} \mathrm{e}^{\mathrm{i}m\phi}\mathrm{e}^{-\mathrm{i}m'\phi'} V^{\mathrm{D}}_{+-;+-}(\boldsymbol{Q}; \boldsymbol{k}_\phi, \boldsymbol{k}'_{\phi'}) \\
&= \mathrm{e}^{\mathrm{i}(m-m')\pi} \int_{\phi,\phi'} \mathrm{e}^{\mathrm{i}m(\phi-\pi)}\mathrm{e}^{-\mathrm{i}m'(\phi'-\pi)} V^{\mathrm{D}}_{+-;+-}(\boldsymbol{Q}; \boldsymbol{k}_\phi, \boldsymbol{k}'_{\phi'})^* \\
&= (-1)^{m-m'} \int_{\phi,\phi'} \mathrm{e}^{\mathrm{i}m\phi}\mathrm{e}^{-\mathrm{i}m'\phi'} V^{\mathrm{D}}_{+-;+-}(\boldsymbol{Q}; \boldsymbol{k}_{\phi+\pi}, \boldsymbol{k}'_{\phi'+\pi}) \\
&= (-1)^{m-m'} \left[\int_{\phi,\phi'} \mathrm{e}^{-\mathrm{i}m\phi}\mathrm{e}^{\mathrm{i}m'\phi'} V^{\mathrm{D}}_{+-;+-}(\boldsymbol{Q}; \boldsymbol{k}_\phi, \boldsymbol{k}'_{\phi'}) \right]^* \\
&= (-1)^{m-m'} \mathcal{V}^{(m,m')}_{+-;+-}(\boldsymbol{Q}; k, k')^*. \tag{3.71}
\end{aligned}$$

We now move on to the subspace with $s = t$. It holds that

$$V^{\mathrm{D/X}}_{++;++}(Q\hat{\boldsymbol{x}}; \boldsymbol{k}_\phi, \boldsymbol{k}'_{\phi'}) = V^{\mathrm{D/X}}_{++;++}(Q\hat{\boldsymbol{x}}; \boldsymbol{k}_{-\phi}, \boldsymbol{k}'_{-\phi'})^*, \tag{3.72a}$$

$$V^{\mathrm{X}}_{++;--}(Q\hat{\boldsymbol{x}}; \boldsymbol{k}_\phi, \boldsymbol{k}'_{\phi'}) = V^{\mathrm{X}}_{++;--}(Q\hat{\boldsymbol{x}}; \boldsymbol{k}_{-\phi}, \boldsymbol{k}'_{-\phi'})^*, \tag{3.72b}$$

$$V^{\mathrm{D}}_{++;++}(\boldsymbol{Q}; \boldsymbol{k}_\phi, \boldsymbol{k}'_{\phi'}) = V^{\mathrm{D}}_{++;++}(\boldsymbol{Q}; \boldsymbol{k}_{\phi+\pi}, \boldsymbol{k}'_{\phi'+\pi}), \tag{3.72c}$$

$$V^{\mathrm{X}}_{++;--}(\boldsymbol{Q}; \boldsymbol{k}_\phi, \boldsymbol{k}'_{\phi'}) = V^{\mathrm{X}}_{++;--}(\boldsymbol{Q}; \boldsymbol{k}_{\phi+\pi}, \boldsymbol{k}'_{\phi'}) = V^{\mathrm{X}}_{++;--}(\boldsymbol{Q}; \boldsymbol{k}_\phi, \boldsymbol{k}'_{\phi'+\pi}). \tag{3.72d}$$

Note that (3.72c) does not come with a complex conjugate such as in Eq. (3.69b). The two first relations again imply that the potential matrix is real when $\boldsymbol{Q}$ is chosen along the x-axis. Via a similar procedure as before, the third one yields

$$\mathcal{V}^{(m,m')}_{++;++}(\boldsymbol{Q}; k, k') = (-1)^{m-m'} \mathcal{V}^{(m,m')}_{++;++}(\boldsymbol{Q}, k, k'). \tag{3.73}$$

This implies that this subblock vanishes identically when $m - m'$ is odd, i.e., when m and m' have opposite parities. Note, however, that the kinetic matrix still couples subspaces of angular momenta with different parities. Similarly, Eq. (3.72d) implies that

$$\mathcal{V}^{(m,m')}_{++;--}(\boldsymbol{Q}; k, k') = (-1)^m \mathcal{V}^{(m,m')}_{++;--}(\boldsymbol{Q}, k, k') = (-1)^{m'} \mathcal{V}^{(m,m')}_{++;--}(\boldsymbol{Q}, k, k'), \tag{3.74}$$

which means that this element vanishes when either m or m' is odd. Furthermore, the sectors with opposite spin-orbit parities are related through

$$
\begin{aligned}
\mathcal{V}^{(m,m')}_{++;++}(Q;k,k') &= \int_{\phi,\phi'} e^{-im\phi} e^{im'\phi'} (V^{\mathrm{D}} - V^{\mathrm{X}})_{++;++}(Q;k_\phi, k'_{\phi'}) \\
&= \left[\int_{\phi,\phi'} e^{im\phi} e^{-im'\phi'} (V^{\mathrm{D}} - V^{\mathrm{X}})_{--;--}(Q;k_\phi, k'_{\phi'}) \right]^* \\
&= \mathcal{V}^{(-m,-m')}_{--;--}(Q;k,k')^*.
\end{aligned}
\tag{3.75}
$$

Finally, using the Hermiticity property we can also find

$$
\begin{aligned}
\mathcal{V}^{(-m,-m')}_{++;--}(Q;k,k') &= -\int_{\phi,\phi'} e^{im\phi} e^{-im'\phi'} V^{\mathrm{X}}_{++;--}(Q;k_\phi, k'_{\phi'}) \\
&= -\int_{\phi,\phi'} e^{im\phi} e^{-im'\phi'} V^{\mathrm{X}}_{--;++}(Q;k'_{\phi'}, k_\phi)^* \\
&= -\int_{\phi,\phi'} e^{im\phi} e^{-im'\phi'} V^{\mathrm{X}}_{++;--}(Q;k'_{\phi'}, k_\phi) \\
&= -\int_{\phi,\phi'} e^{-im'\phi} e^{im\phi'} V^{\mathrm{X}}_{++;--}(Q;k'_\phi, k_{\phi'}) \\
&= \mathcal{V}^{(m',m)}_{++;--}(Q;k',k).
\end{aligned}
\tag{3.76}
$$

3.D Analysis of the exciton problem with $\phi_Q \neq 0$

In this appendix we carefully study the exciton problem for the general case when Q forms a nonzero angle ϕ_Q with the x-axis. This is needed to determine Berry curvatures and other topological properties, as they depend on the entire Q-space. By rotational symmetry, when $\phi_Q \neq 0$ the exciton wave functions have to be related to the wave functions at $\phi_Q = 0$ via

$$
\Phi_{Q;s,t}(k,\phi) = e^{i\varphi_{Q;s,t}} \Phi_{Q\hat{x};s,t}(k,\phi - \phi_Q),
\tag{3.77}
$$

that is, they can at most differ by a global phase. In this section it is important to remember the property

$$
V^{\mathrm{D/X}}_{s,t;s',t'}(Q;k,k') = V^{\mathrm{D/X}}_{-s,-t;-s',-t'}(Q;k,k')^*.
\tag{3.78}
$$

We start with the simpler case of the two subspaces with $s \neq t$, which are independent. For this reason, and because of the property (3.78), we can always choose $\Phi_{Q;+-}(k) = \Phi^*_{Q;-+}(k)$ for arbitrary Q. Furthermore, the nonzero potentials in this subspace satisfy

$$
V^{\mathrm{D}}_{+-;+-}(Q;k_\phi, k'_{\phi'}) = V^{\mathrm{D}}_{+-;+-}(Q\hat{x};k_{\phi-\phi_Q}, k'_{\phi'-\phi_Q}).
\tag{3.79}
$$

This further implies that we may always choose $\varphi_{Q;+-} = \varphi_{Q;-+} = 0$, as the exciton eigenvalue problem in the variables (Q, k_ϕ) is the same as in the variables $(Q\hat{x}, k_{\phi-\phi_Q})$. In summary, we can always choose a gauge such that

$$
\Phi_{Q;+-}(k,\phi) = \Phi^*_{Q;-+}(k,\phi) = \Phi_{Q\hat{x};+-}(k,\phi - \phi_Q).
\tag{3.80}
$$

In view of the degeneracy of these two subspaces, we can write down two doublet eigenstates

$$
|Q;0_\pm\rangle = \frac{1}{\sqrt{2}} \Big(|Q;+,-\rangle \pm |Q;-,+\rangle \Big),
\tag{3.81}
$$

with the convention that the wave functions contained in $|Q; +, -\rangle$ and $|Q; -, +\rangle$ satisfy (3.80).

On the other hand, the subspaces with $s = t$ are coupled by the exchange interaction for nonzero Q. The property (3.78) imposes the following constraint between the two wave functions:

$$\Phi_{Q;--}(k) = e^{i\xi_Q} \Phi_{Q;++}^*(k). \tag{3.82}$$

However, we cannot simply set this phase factor to unity due to the coupling. To determine ξ_Q we must take into account the following additional properties relating the potentials for nonzero ϕ_Q with those for $\phi_Q = 0$:

$$V_{++;++}^{D/X}(Q; k_\phi, k'_{\phi'}) = V_{++;++}^{D/X}(Q\hat{x}; k_{\phi-\phi_Q}, k'_{\phi'-\phi_Q}), \tag{3.83a}$$

$$V_{++;--}^{X}(Q; k_\phi, k'_{\phi'}) = e^{-2i\phi_Q} V_{++;--}^{X}(Q\hat{x}; k_{\phi-\phi_Q}, k'_{\phi'-\phi_Q}). \tag{3.83b}$$

These, together with (3.77) as well as the requirement that the functions $\Phi_{Q\hat{x};s,t}^{(m)}(k)$ be real, imply the constraints

$$e^{i(\xi_Q - \varphi_{Q;++} - \varphi_{Q;--})} = \pm 1, \tag{3.84a}$$

$$e^{i(\varphi_{Q;++} - \varphi_{Q;--} + 2\phi_Q)} = 1. \tag{3.84b}$$

The two signs in the first equation correspond to the wave functions for the two states that evolve from each degenerate pair in the families $|0; +, +\rangle$ and $|0; -, -\rangle$. However, note that the wave functions in the state with the upper sign are generally *not* equal to those in the state with the lower sign. There are no further requirements, because with these constraints $\varphi_{Q;++}$ becomes a global phase of the eigenstate:

$$\begin{bmatrix} \Phi_{Q;++}(k, \phi) \\ \Phi_{Q;--}(k, \phi) \end{bmatrix} = e^{i\varphi_{Q;++}} \begin{bmatrix} \Phi_{Q\hat{x};++}(k, \phi - \phi_Q) \\ \pm e^{2i\phi_Q} \Phi_{Q\hat{x};++}^*(k, \phi - \phi_Q) \end{bmatrix}. \tag{3.85}$$

In view of the above equation, we now choose $e^{i\varphi_{Q;++}} = e^{-i\phi_Q}$ and write the eigenstates of the problem as

$$|Q; 2_\pm\rangle = \frac{1}{\sqrt{2}} \left(e^{-i\phi_Q} |Q; +, +\rangle \pm e^{i\phi_Q} |Q; -, -\rangle \right). \tag{3.86}$$

Note that the choice of $\varphi_{Q;++}$ does not affect the pseudospin winding number. To write the eigenstates in this form, we have explicitly separated the phase factors and again adopted the convention that the wave functions contained in $|Q; +, +\rangle$ and $|Q; -, -\rangle$ satisfy

$$\Phi_{Q;++}(k, \phi) = \Phi_{Q;--}^*(k, \phi) = \Phi_{Q\hat{x};++}(k, \phi - \phi_Q). \tag{3.87}$$

3.E Derivation of the effective model for small Q

Here we derive the effective model of Eq. (3.46) by applying degenerate perturbation theory around $Q = 0$ in the subspace with $s = t$. We want to analyze the behavior of the states $|Q; +, +; n, m\rangle$ and $|Q; -, -; n, -m\rangle$ for small total momentum, as they are degenerate at $Q = 0$.

We need to compute the matrix elements $\langle Q; s, s; n, m|\hat{H}|Q; s', s'; n, -m\rangle$, which in principle should be done by using the wave functions obtained from solving the exciton problem at nonzero Q without exchange interaction. However, for small Q, we can in first approximation use the wave functions at $Q = 0$. Some care must be taken when doing this, because even though $|Q|$ is small, the angle ϕ_Q must be taken to be fixed and arbitrary. Then, we can write

$$\Phi_{Q;s,t}(k)\Big|_{Q\to 0} \approx \Phi_{0;s,t}^{(m)}(k) e^{im(\phi_k - \phi_Q)}, \tag{3.88}$$

where we have only kept the dominant Fourier coefficient in the zero-momentum limit, which is real. As always, we have omitted the principal quantum number. Note that the phase $e^{i\varphi_{Q;s,t}}$ of the previous section is chosen as unity to respect the rotational symmetry of the problem. Then we have, for instance,

$$\langle Q; +, +; n, m | \hat{V}^{D/X} | Q; +, +; n, m \rangle \approx \int_0^\infty \frac{k \mathrm{d}k}{(2\pi)^2} \int_0^\infty \frac{k' \mathrm{d}k'}{(2\pi)^2} \, \Phi_{0;++}^{(m)}(k) \, \Phi_{0;++}^{(m)}(k') \qquad (3.89)$$

$$\times \int_0^{2\pi} \mathrm{d}\phi \int_0^{2\pi} \mathrm{d}\phi' \, e^{-im(\phi - \phi')} \, V_{s,t;s',t'}^{D/X}(Q; k_\phi, k'_{\phi'}).$$

If we split the Hamiltonian as $\hat{H} = \hat{K} + \hat{V}^D - \hat{V}^X$, where $\hat{K}$ contains the band structure, then under this aproximation one finds

$$\langle Q; +, +; n, m | \hat{K} + \hat{V}^D | Q; +, +; n, m \rangle \approx \omega_{n,m} + \frac{Q^2}{2M_{n,m}} + \mathcal{O}(Q^4). \qquad (3.90)$$

Here, $\omega_{n,m}$ is the energy of the exciton at $Q = 0$ and $M_{n,m}$ is an effective mass. Both of these parameters depend on m and the principal quantum number n labeling the different excited states. On the other hand, using the property

$$V_{++;++}^X(Q\hat{x}; k_\phi, k'_{\phi'}) = V_{++;--}^X(Q\hat{x}; k_\phi, k'_{-\phi'}), \qquad (3.91)$$

it is straightforward to show that

$$\langle Q; +, +; n, m | \hat{V}^X | Q; -, -; n, -m \rangle \approx e^{-2i\phi_Q} \langle Q; +, +; n, m | \hat{V}^X | Q; +, +; n, m \rangle. \qquad (3.92)$$

Putting everything together, we arrive at the effective Hamiltonian

$$H_{n,m}^{\mathrm{eff}}(Q) = \left(\omega_{n,m} + \frac{Q^2}{2M_{n,m}} \right) \mathbb{I}_2 + \mathcal{J}_{n,m}(Q) \begin{bmatrix} 1 & e^{-2i\phi_Q} \\ e^{2i\phi_Q} & 1 \end{bmatrix}. \qquad (3.93)$$

Below we show that the coupling $\mathcal{J}_{n,m}(Q)$ reads as follows:

$$\mathcal{J}_{n,m}(Q) = \begin{cases} 0 & \text{if } m \text{ is odd}, \\ J_{n,m} |Q| + \dfrac{Q^2}{2M'_{n,m}} + \mathcal{O}(|Q|^3) & \text{if } m \text{ is even}, \end{cases} \qquad (3.94)$$

and we expect that $J_{n,m\neq 0} \ll J_{n,0}$. It should be stressed that the excitons with odd angular momentum only remain degenerate in our perturbative treatment where we consider the wave functions at $Q = 0$. When higher-order Q terms are considered in the wave functions, the degeneracy is lifted for these excitons as well. However, this effect is of order $|Q|^3$.

To see that Eq. (3.94) holds we start from the expression for $\mathcal{J}_{n,m}(Q)$, which reads

$$\mathcal{J}_{n,m}(Q) = -V(Q) \left| \frac{1}{V} \sum_k \langle \chi_{-Q/2+k}^{v,+} | \chi_{Q/2+k}^{c,+} \rangle \Phi_{0;++}^{(m)}(k) \, e^{im\phi_k} \right|^2. \qquad (3.95)$$

Since this is a real quantity, it should be independent of ϕ_Q, because the exciton spectrum can only depend on the magnitude of the total momentum. To see that this is indeed the case, we note that

$$\langle \chi_{-Q/2+k_\phi}^{v,+} | \chi_{Q/2+k_\phi}^{c,+} \rangle = e^{i\phi_Q} \langle \chi_{-Q\hat{x}/2+k_{\phi-\phi_Q}}^{v,+} | \chi_{Q\hat{x}/2+k_{\phi-\phi_Q}}^{c,+} \rangle. \qquad (3.96)$$

Therefore, after a change of variable, the integrand in Eq. (3.95) only depends on ϕ_Q through a global phase. We can expand the inner product in powers of $|Q|$ by recalling that it vanishes linearly when $|Q| \to 0$:

$$\langle \chi_{-Q/2+k}^{v,+} | \chi_{Q/2+k}^{c,+} \rangle = \alpha(k) |Q| + \frac{1}{2}\beta(k) Q^2 + \mathcal{O}(|Q|^3). \qquad (3.97)$$

71

Note that the ϕ_Q-dependence of the coefficients has been omitted in view of our comment above. Assuming that the coefficient $\alpha(k)$ only varies slowly with k yields

$$\mathcal{J}_{n,m}(Q) = -V(Q)\left|\alpha(0)\,\Phi_{0;++}^{n,m}(r=0)|Q| + \mathcal{O}(Q^2)\right|^2. \tag{3.98}$$

Here, $\Phi_{0;++}^{n,m}(r=0)$ is the real-space probability amplitude at the origin, which is nonzero only for s-wave excitons. Thus, renaming the couplings, in this case we obtain

$$\mathcal{J}_{n,0}(Q) = J_{n,0}|Q| + \frac{Q^2}{2M'_{n,0}} + \mathcal{O}(|Q|^3), \tag{3.99}$$

where we have used that $V(Q) \propto |Q|^{-1}$ in the long-wavelength limit. For even-m excitons with $m \neq 0$ a similar expression will hold, but the linear coupling is expected to be much weaker due to the fact that the wave function is zero at the origin. [The coupling is not necessarily zero because α in Eq. (3.97) is still a function of k, hence higher-order k terms can contribute.] For odd-m excitons the entire expression of Eq. (3.95) vanishes, which follows from our discussion in App. 3.C. However, as explained above, the degeneracy is broken to third order in $|Q|$ when including higher-order terms in the wave functions themselves. This escapes our perturbative treatment, which is therefore valid to order Q^2.

4

Many-body effects

In the previous chapter we have solved the exciton problem in Bi_2Se_3 in the simplest case. That is, we have neglected the influence of many-body effects, such as interactions with the topological electronic surface states, and obtained an idealized picture of bulk excitons. The aim of this chapter is to refine this picture by including these effects in a perturbative scheme. Our main interest is to find out how the surface states screen the bare bulk interaction and to understand the new physics that arise from this screening. One important thing to note is that we shall only be concerned with corrections to the two-body electron-hole potential. The analysis of higher-body interactions is out of the scope of this thesis, and hence we will systematically ignore the corresponding terms. Nevertheless, our methods and procedure can be straightforwardly generalized to include these effects as well, if so desired.

4.1 Path integral setup

Many-body effects are most conveniently studied by means of the path integral formalism. In this section we set up its essential ingredients and introduce some convenient notation to be used in the remainder of this chapter. Once again we consider a thin slab of thickness ℓ located on the xy-plane, and start from the full three-dimensional model as described by the Hamiltonian (2.7). Simplifications such as integrating out z-dependences will be performed as we proceed.

We firstly introduce a set of fermionic (i.e., Grassmann-valued) bulk and surface fields, which we denote by ϕ_B and ϕ_S, respectively. As in Ch. 2, we assume that these can be treated independently due to the robustness of the surface states. In first instance, these fields depend on the position (x, z) in real space, on the imaginary time τ, and on the orbital and spin degrees of freedom $i \in \{Bi^+, Se^-\}$ and $\sigma \in \{\uparrow, \downarrow\}$, respectively. To shorten the notation we combine the latter into an index $\mathbf{i} \equiv (i, \sigma)$. The partition function reads [90, Ch. 7, 113, Ch. 4]

$$\mathcal{Z} = \int \mathcal{D}[\phi^*]\mathcal{D}[\phi] \, \exp(-S[\phi^*, \phi]), \tag{4.1}$$

where our total action $S[\phi^*, \phi]$ is defined as

$$S = S_{B,0} + S_{S,0} + S_{int}. \tag{4.2}$$

The free term $S_{B,0}$ represents the free bulk action and is given by

$$S_{B,0}[\phi_B^*, \phi_B] = \sum_{\mathbf{i},\mathbf{i}'} \int d^2x \int dz \int d\tau\, \phi_{B;\mathbf{i}}^*(\boldsymbol{x}, z, \tau) \left[\delta_{\mathbf{i}\mathbf{i}'} \frac{\partial}{\partial \tau} + H(-i\boldsymbol{\nabla}, -i\partial_z)_{\mathbf{i}\mathbf{i}'} - \mu\delta_{\mathbf{i}\mathbf{i}'} \right] \phi_{B;\mathbf{i}'}(\boldsymbol{x}, z, \tau).$$

(4.3)

Similarly, $S_{S,0}$ represents the free surface action and is obtained from the above equation by simply changing ϕ_B to ϕ_S. Here, $H(-i\boldsymbol{\nabla}, -i\partial_z)$ stands for the full Bi$_2$Se$_3$ Hamiltonian (2.7) after formally performing the substitution $k_i \rightarrow -i\partial_i$, and μ is the chemical potential. In this chapter, all integrals over z are implicitly restricted to the interval $[-\ell/2, \ell/2]$, and those over τ run from 0 to $\beta = (k_B T)^{-1}$.

On the other hand, the interacting part of the action is

$$S_{int}[\phi^*, \phi] = \frac{1}{2} \sum_{\mathbf{i},\mathbf{i}'} \sum_{A,A'} \int d^2x \int d^2x' \int dz \int dz' \int d\tau\, \phi_{A;\mathbf{i}}^*(\boldsymbol{x}, z, \tau)\, \phi_{A';\mathbf{i}'}^*(\boldsymbol{x}', z', \tau)$$
$$\times V(\boldsymbol{x} - \boldsymbol{x}'; z, z')\, \phi_{A';\mathbf{i}'}(\boldsymbol{x}', z', \tau)\, \phi_{A;\mathbf{i}}(\boldsymbol{x}, z, \tau),$$

(4.4)

where $A, A' \in \{B, S\}$ and V is the bare interaction potential between electrons in the sample, assumed to be instantaneous in imaginary time. Here we have assumed that the robustness of the surfaces remains intact even when switching on the interaction between particles. In other words, a coherent surface state is still forbidden to scatter into the bulk, so there are no terms of the form $\phi_B^* \phi_B^* \phi_B \phi_S$, for instance. We note that V is determined from electrostatics and is thus independent of the microscopic details, which means it cannot change the orbital or spin quantum numbers. Translational symmetry in the xy-plane ensures that V only depends on the in-plane coordinates $\boldsymbol{x}$ and $\boldsymbol{x}'$ through their difference. By contrast, the thin slab geometry may give a more complicated dependence on the z and z' coordinates as long as it is still symmetric under the exchange of z and z', such as in the case of the Rytova-Keldysh potential. Since the details of the interaction potential are at present unimportant, we leave it unspecified for the time being.

We now expand the surface fields in the basis of surface states (2.19) by writing

$$\phi_{S;\mathbf{i}}(\boldsymbol{x}, z, \tau) = \sum_{\mathbf{s}} \langle \mathbf{i} | \psi_{\mathbf{s}}(z) \rangle\, \phi_{S;\mathbf{s}}(\boldsymbol{x}, \tau).$$

(4.5)

Here, $\mathbf{s} \equiv (s, \alpha)$ labels the different basis states, with $s \in \{t, b\}$ and $\alpha \in \{\uparrow, \downarrow\}$. Note that $\langle \mathbf{i} | \psi_{\mathbf{s}}(z) \rangle$ stands for the component-wise inner product and in particular selects the $\mathbf{i}$ component of the state $\psi_{\mathbf{s}}(z)$. For completeness, the inverse transformation is

$$\phi_{S;\mathbf{s}}(\boldsymbol{x}, \tau) = \sum_{\mathbf{i}} \int dz\, \langle \psi_{\mathbf{s}}(z) | \mathbf{i} \rangle\, \phi_{S;\mathbf{i}}(\boldsymbol{x}, z, \tau).$$

(4.6)

This expansion allows us to work explicitly with the fields $\phi_{S;\mathbf{s}}$ describing the two individual surfaces themselves, which is desirable from a physical point of view. Furthermore, it allows us to formally integrate out all the z-dependences and obtain a two-dimensional effective model. The free part of the action $S_{S,0}$ becomes

$$S_{S,0}[\phi_S^*, \phi_S] = \sum_{\mathbf{s},\mathbf{s}'} \int d^2x \int d\tau\, \phi_{S;\mathbf{s}}^*(\boldsymbol{x}, \tau) \left[\delta_{\mathbf{s}\mathbf{s}'} + H_S(-i\boldsymbol{\nabla})_{\mathbf{s}\mathbf{s}'} - \mu\delta_{\mathbf{s}\mathbf{s}'} \right] \phi_{S;\mathbf{s}'}(\boldsymbol{x}, \tau),$$

(4.7)

where we have identified the components (2.21) of the effective surface Hamiltonian. For convenience we remind the reader that it is given by

$$H_S(\boldsymbol{k}) = \begin{bmatrix} h_S(\boldsymbol{k}) & 0 \\ 0 & h_S(\boldsymbol{k}) \end{bmatrix} \quad \text{with} \quad h_S(\boldsymbol{k}) = E_0 + D|\boldsymbol{k}|^2 - v_F(\boldsymbol{k} \times \boldsymbol{\sigma}) \cdot \hat{\boldsymbol{z}}.$$

(4.8)

On the other hand, the part of the action that describes the surface-surface interactions now becomes

$$S_{\text{int}}^{(\text{SS})}[\phi_S^*, \phi_S] = \frac{1}{2} \sum_{s,s'} \int d^2x \int d^2x' \int d\tau \, \phi_{S;s}^*(\boldsymbol{x}, \tau) \, \phi_{S;s'}^*(\boldsymbol{x}', \tau)$$
$$\times V_{ss'}^{\text{qc}}(\boldsymbol{x} - \boldsymbol{x}') \, \phi_{S;s'}(\boldsymbol{x}', \tau) \, \phi_{S;s}(\boldsymbol{x}, \tau). \tag{4.9}$$

In this equation we have identified a quantum-confined version of the interaction potential in a manner similar to Refs. [97, 114], which reads

$$V_{ss'}^{\text{qc}}(\boldsymbol{x} - \boldsymbol{x}') = \int dz \int dz' \, \pi_s(z) \, V(\boldsymbol{x} - \boldsymbol{x}'; z, z') \, \pi_{s'}(z'). \tag{4.10}$$

Here, $\pi_s(z) \equiv \psi_s^\dagger(z) \psi_s(z)$ is the spin-independent probability density of the surface s, obtained from the eigenstates (2.19). Note that the orthonormality property (2.20) is crucial to obtain this result.

For convenience, we now introduce the notation from Ref. [90] to simultaneously denote discrete and continuous matrix products by defining

$$(\Phi|M|\Psi) \equiv \sum_{x,x'} \Phi^*(x) \, M(x, x') \, \Psi(x'). \tag{4.11}$$

Here, x and x' stand for all indices and coordinates the fields Φ and Ψ depend on, which may be discrete, continuous (in which case the sum is understood to be an integral), or a combination of both. Then we can simply rewrite the full action as

$$S = \underbrace{(\phi_B|-G_{B,0}^{-1}|\phi_B)}_{S_{B,0}} + \underbrace{(\phi_S|-G_{S,0}^{-1}|\phi_S)}_{S_{S,0}} + \underbrace{\frac{1}{2} \sum_{A,A'} (\phi_A^* \phi_A |V| \phi_{A'}^* \phi_{A'})}_{S_{\text{int}}}, \tag{4.12}$$

where the free (inverse) bulk and surface Green's functions are identified as

$$G_{B,0}^{-1}(\boldsymbol{x}, z, \tau; \boldsymbol{x}', z', \tau') = -\left[\frac{\partial}{\partial \tau} + H(-i\boldsymbol{\nabla}, -i\partial_z) - \mu \right] \delta(\boldsymbol{x} - \boldsymbol{x}') \, \delta(z - z') \, \delta(\tau - \tau'), \tag{4.13}$$

$$G_{S,0}^{-1}(\boldsymbol{x}, \tau; \boldsymbol{x}', \tau') = -\left[\frac{\partial}{\partial \tau} + H_S(-i\boldsymbol{\nabla}) - \mu \right] \delta(\boldsymbol{x} - \boldsymbol{x}') \, \delta(\tau - \tau'). \tag{4.14}$$

Note that in writing the interaction term in Eq. (4.12) the interaction matrix V is understood to be $V(\boldsymbol{x}, z, \tau; \boldsymbol{x}', z', \tau') = V(\boldsymbol{x} - \boldsymbol{x}', z, z') \delta(\tau - \tau')$.

4.2 Effective bulk action

In this section we will integrate out the fermionic surface fields and obtain an effective action for the bulk fields. The first step is to decouple the term $(\phi_S^* \phi_S |V| \phi_S^* \phi_S)$, which is not quadratic in the surface fields and thus prevents us from integrating them out exactly. We perform a Hubbard-Stratonovich transformation by introducing two real fields $\rho_s(\boldsymbol{x}, \tau)$ with $s \in \{t, b\}$, which will represent the total densities on each surface. We write

$$1 = \int \mathcal{D}[\rho] \exp\left\{ \frac{1}{2} \sum_{s,s'} \int d^2x \int d^2x' \int d\tau \, \rho_s(\boldsymbol{x}, \tau) \, V_{ss'}^{\text{qc}}(\boldsymbol{x} - \boldsymbol{x}') \, \rho_{s'}(\boldsymbol{x}', \tau) \right\}, \tag{4.15}$$

where the measure is understood to contain the factor of $\exp\{\mathrm{Tr}[\log(-V_{\mathrm{qc}}^{-1})]/2\}$ coming from the path integral so that the whole expression equals unity. We now perform a shift in the density fields by replacing

$$\rho_s(\boldsymbol{x},\tau) \to \rho_s(\boldsymbol{x},\tau) - \sum_\alpha \phi_{\mathrm{S};s,\alpha}^*(\boldsymbol{x},\tau)\,\phi_{\mathrm{S};s,\alpha}(\boldsymbol{x},\tau). \tag{4.16}$$

Multiplying the partition function by this path integral over the density fields then gets rid of the term in Eq. (4.9).

Let us now analyze the surface-bulk interaction term, $S_{\mathrm{int}}^{(\mathrm{SB})} = (\phi_{\mathrm{S}}^*\phi_{\mathrm{S}}|V|\phi_{\mathrm{B}}^*\phi_{\mathrm{B}})$. It is quadratic in the surface fields, so in principle it poses no problem to integrate over them. Indeed, we may rewrite it as a self-energy-like term $(\phi_{\mathrm{S}}|\mathcal{M}_{\mathrm{B}}|\phi_{\mathrm{S}})$, where $\mathcal{M}_{\mathrm{B}}$ is a matrix that depends on the bulk fields:

$$\mathcal{M}_{\mathrm{B};\mathbf{ss'}}(\boldsymbol{x},\tau;\boldsymbol{x'},\tau') = \delta_{\mathbf{ss'}}\,\delta(\boldsymbol{x}-\boldsymbol{x'})\,\delta(\tau-\tau')\sum_{\mathbf{i''}}\int \mathrm{d}^2x'' \int \mathrm{d}z \int \mathrm{d}z''$$
$$\times\, \pi_s(z)\,V(\boldsymbol{x}-\boldsymbol{x''};z,z'')\,\phi_{\mathrm{B};\mathbf{i''}}^*(\boldsymbol{x''},z'',\tau)\,\phi_{\mathrm{B};\mathbf{i''}}(\boldsymbol{x''},z'',\tau). \tag{4.17}$$

At this point the total action, with the additional density fields included, reads

$$S = (\phi_{\mathrm{B}}|-G_{\mathrm{B},0}^{-1}|\phi_{\mathrm{B}}) + (\phi_{\mathrm{S}}|-G_{\mathrm{S},0}^{-1} + \mathcal{M}_{\mathrm{B}}|\phi_{\mathrm{S}}) + \frac{1}{2}(\phi_{\mathrm{B}}^*\phi_{\mathrm{B}}|V|\phi_{\mathrm{B}}^*\phi_{\mathrm{B}}) - \frac{1}{2}(\rho|V^{\mathrm{qc}}|\rho) + (\rho|V^{\mathrm{qc}}|\phi_{\mathrm{S}}^*\phi_{\mathrm{S}}). \tag{4.18}$$

The last term may be rewritten as $(\phi_{\mathrm{S}}|\Sigma_{\mathrm{S}}[\rho]|\phi_{\mathrm{S}})$, where

$$\Sigma_{\mathrm{S}}[\rho]_{\mathbf{ss'}}(\boldsymbol{x},\tau;\boldsymbol{x'},\tau') = \delta_{\mathbf{ss'}}\,\delta(\boldsymbol{x}-\boldsymbol{x'})\,\delta(\tau-\tau')\sum_{s''}\int \mathrm{d}^2x''\, V_{ss''}^{\mathrm{qc}}(\boldsymbol{x}-\boldsymbol{x''})\,\rho_{s''}(\boldsymbol{x''},\tau). \tag{4.19}$$

Then the action is quadratic in the surface fields and depends on them only through a term $(\phi_{\mathrm{S}}|-G_{\mathrm{S}}[\rho]^{-1} + \mathcal{M}_{\mathrm{B}}|\phi_{\mathrm{S}})$, where we have defined

$$G_{\mathrm{S}}[\rho]^{-1} \equiv G_{\mathrm{S},0}^{-1} - \Sigma_{\mathrm{S}}[\rho]. \tag{4.20}$$

Integrating out this term gives the standard factor of $\exp[\mathrm{Tr}\log(-G_{\mathrm{S}}[\rho]^{-1} + \mathcal{M}_{\mathrm{B}})]$. Hence, we are left with

$$S = (\phi_{\mathrm{B}}|-G_{\mathrm{B},0}^{-1}|\phi_{\mathrm{B}}) + \frac{1}{2}(\phi_{\mathrm{B}}^*\phi_{\mathrm{B}}|V|\phi_{\mathrm{B}}^*\phi_{\mathrm{B}}) - \frac{1}{2}(\rho|V^{\mathrm{qc}}|\rho) - \mathrm{Tr}\log(-G_{\mathrm{S}}[\rho]^{-1} + \mathcal{M}_{\mathrm{B}}). \tag{4.21}$$

We now concentrate on the last two terms. We perform a fluctuation expansion of the density fields around their minimum by writing $\rho_s(\boldsymbol{x},\tau) = \rho_{0;s}(\boldsymbol{x},\tau) + \delta\rho_s(\boldsymbol{x},\tau)$, such that $\rho_{0;s}(\boldsymbol{x},\tau) = \langle\rho_s(\boldsymbol{x},\tau)\rangle$. Then we can write

$$\log(-G_{\mathrm{S}}[\rho]^{-1} + \mathcal{M}_{\mathrm{B}}) = \log(-G_{\mathrm{S}}[\rho_0]^{-1}) - \sum_{n=1}^{\infty}\frac{1}{n}\big(G_{\mathrm{S}}[\rho_0]\Sigma_{\mathrm{S}}[\delta\rho] + G_{\mathrm{S}}[\rho_0]\mathcal{M}_{\mathrm{B}}\big)^n. \tag{4.22}$$

Demanding that the noninteracting terms linear in the density fluctuations vanish gives the condition $\mathrm{Tr}\big(G_{\mathrm{S}}[\rho_0]\Sigma_{\mathrm{S}}[\delta\rho]\big) - (\rho_0|V^{\mathrm{qc}}|\delta\rho) = 0$. The fact that $\delta\rho$ is arbitrary leads to

$$\langle\rho_s(\boldsymbol{x},\tau)\rangle = \sum_\alpha G_{\mathrm{S}}[\rho_0]_{s,\alpha;s,\alpha}(\boldsymbol{x},\tau;\boldsymbol{x},\tau), \tag{4.23}$$

which implicitly defines the mean surface densities $\rho_{0;s}$.

The rest of the terms in the summation of Eq. (4.22) can be conveniently represented as Feynman diagrams. After dropping some unimportant constant terms that only depend on ρ_0, the action is now

$$S = \left(\phi_{\mathrm{B}}\middle|-G_{\mathrm{B},0}^{-1}\middle|\phi_{\mathrm{B}}\right) + \frac{1}{2}\left(\phi_{\mathrm{B}}^*\phi_{\mathrm{B}}\middle|V\middle|\phi_{\mathrm{B}}^*\phi_{\mathrm{B}}\right) - \frac{1}{2}\left(\delta\rho\middle|V^{\mathrm{qc}}\middle|\delta\rho\right)$$

$$\tag{4.24}$$

where the external solid lines represent bulk fermions, the dashed lines represent the surface propagator $G_{\mathrm{S}}[\rho_0]$, and the crossed dots stand for the surface-density fluctuations $\delta\rho$. The notation $\int\{\dots\}$ implies that we sum and integrate over the coordinates of all free ends. The series is understood to contain all possible diagrams consisting of a single surface loop with an arbitrary number and ordering of attached external legs of both types [except for the diagram representing $\mathrm{Tr}(G_{\mathrm{S}}[\rho_0]\Sigma_S[\delta\rho])$, which has canceled the other noninteracting term linear in $\delta\rho$]. The precise Feynman rules for these kinds of diagrams are given in App. 4.A.

The first diagram in Eq. (4.24) corresponds to the term $G_{\mathrm{S}}[\rho_0]\mathcal{M}_{\mathrm{B}}$ in the expansion of (4.22). It gives a Hartree correction to the free bulk propagator, such that we can write the sum of both terms as $\left(\phi_{\mathrm{B}}\middle|-G_{\mathrm{B}}[\rho_0]^{-1}\middle|\phi_B\right)$, where we have defined

$$G_{\mathrm{B}}[\rho_0]^{-1} \equiv G_{\mathrm{B},0}^{-1} - \Sigma_{\mathrm{B}}[\rho_0]. \tag{4.25}$$

In this case, the bulk self-energy correction is given by

$$\Sigma_{\mathrm{B}}[\rho_0]_{\mathbf{ii'}}(\boldsymbol{x},z,\tau;\boldsymbol{x'},z',\tau') = \delta_{\mathbf{ii'}}\,\delta(\boldsymbol{x}-\boldsymbol{x'})\,\delta(z-z')\,\delta(\tau-\tau')$$

$$\times \sum_{s''}\int \mathrm{d}^2x''\int \mathrm{d}z''\,\pi_{s''}(z'')\,V(\boldsymbol{x}-\boldsymbol{x''};z,z'')\,\langle\rho_{s''}(\boldsymbol{x''},\tau)\rangle. \tag{4.26}$$

On the other hand, the first diagram on the second line stands for $\frac{1}{2}G_{\mathrm{S}}[\rho_0]\mathcal{M}_{\mathrm{B}}G_{\mathrm{S}}[\rho_0]\mathcal{M}_{\mathrm{B}}$. It provides a correction to the bare bulk-bulk interaction term $\frac{1}{2}\left(\phi_{\mathrm{B}}^*\phi_{\mathrm{B}}\middle|V\middle|\phi_{\mathrm{B}}^*\phi_{\mathrm{B}}\right)$, because we can write its trace as $\frac{1}{2}\left(\phi_{\mathrm{B}}^*\phi_{\mathrm{B}}\middle|V^{(1)}\middle|\phi_{\mathrm{B}}^*\phi_{\mathrm{B}}\right)$, where
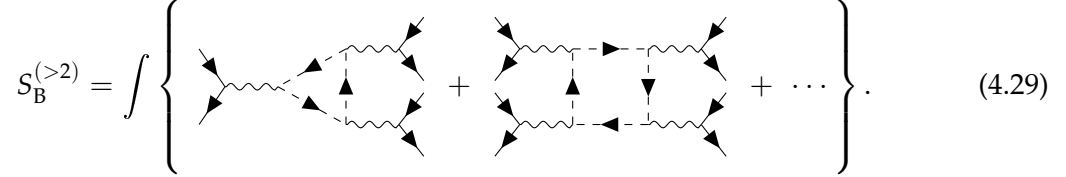
$$V^{(1)}(\boldsymbol{x},z,\tau;\boldsymbol{x'},z',\tau') = \sum_{\mathbf{s},\mathbf{s'}}\int \mathrm{d}^2x''\int \mathrm{d}z''\int \mathrm{d}^2x'''\int \mathrm{d}z'''$$

$$\times \pi_s(z'')\,\pi_{s'}(z''')\,V(\boldsymbol{x}-\boldsymbol{x''},z,z'')\,V(\boldsymbol{x'}-\boldsymbol{x'''},z',z''')$$

$$\times G_{\mathrm{S}}[\rho_0]_{\mathbf{ss'}}(\boldsymbol{x''},\tau;\boldsymbol{x'''},\tau')\,G_{\mathrm{S}}[\rho_0]_{\mathbf{s's}}(\boldsymbol{x'''},\tau';\boldsymbol{x''},\tau). \tag{4.27}$$

We now choose to neglect all terms which are of order $\delta\rho^3$ and higher, that is, all those terms in (4.24) with three or more crossed dots. In this way we obtain an action that is quadratic in the

surface density fields, and hence it is possible to find a vector $\mathcal{R}$ and a matrix $\mathcal{U}$ such that

$$S = (\phi_B | -G_B[\rho_0]^{-1} | \phi_B) + \frac{1}{2}(\phi_B^* \phi_B | V + V^{(1)} | \phi_B^* \phi_B) + S_B^{(>2)} + (\delta\rho | \mathcal{R}) - \frac{1}{2}(\delta\rho | V^{qc} - \mathcal{U} | \delta\rho). \tag{4.28}$$

The term $S_B^{(>2)}$ contains all $\delta\rho$-independent diagrams of order $(\phi_B^* \phi_B)^3$ and higher, that is,

$$S_B^{(>2)} = \int \left\{ \vcenter{\hbox{}} + \vcenter{\hbox{}} + \cdots \right\}. \tag{4.29}$$

Note that $\mathcal{R}$ and $\mathcal{U}$ are determined from the $\delta\rho$-dependent diagrams in (4.24) and that both depend on the bulk fields; in fact, they contain arbitrary powers of $\phi_B^* \phi_B$. Integrating out the bosonic field $\delta\rho$ finally gives an effective bulk action, which reads

$$S_{B,eff}[\phi_B^*, \phi_B] = (\phi_B | -G_B[\rho_0]^{-1} | \phi_B) + \frac{1}{2}(\phi_B^* \phi_B | V + V^{(1)} | \phi_B^* \phi_B) + S_B^{(>2)}[\phi_B^*, \phi_B]$$
$$+ \frac{1}{2}(\mathcal{R} | [V^{qc} - \mathcal{U}]^{-1} | \mathcal{R}) + \frac{1}{2} \operatorname{Tr} \log[-(V^{qc} - \mathcal{U})]. \tag{4.30}$$

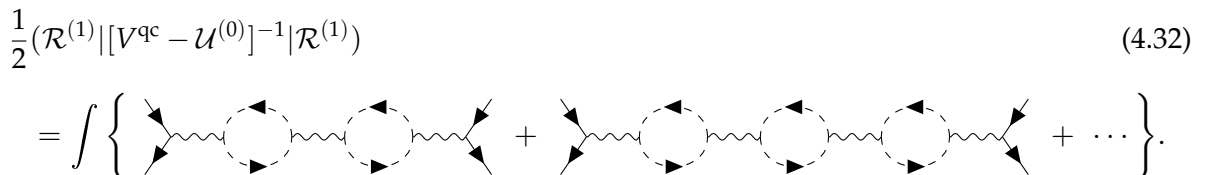The last two terms can be written in a more practical form by using the following identities:

$$(V^{qc} - \mathcal{U})^{-1} = V_{qc}^{-1} + V_{qc}^{-1} \mathcal{U} V_{qc}^{-1} + V_{qc}^{-1} \mathcal{U} V_{qc}^{-1} \mathcal{U} V_{qc}^{-1} + \cdots, \tag{4.31a}$$

$$\log[-(V^{qc} - \mathcal{U})] = \log(-V_{qc}^{-1}) - \sum_{n=1}^{\infty} (V_{qc}^{-1} \mathcal{U})^n. \tag{4.31b}$$

The terms of order $(\phi_B^* \phi_B)^3$ and higher in the effective bulk action provide corrections to the interactions between three or more particles. As mentioned at the outset, however, in this chapter we are only interested in the effects of many-body phenomena on a single exciton, i.e., on the two-body electron-hole problem. The Feynman diagrams contained in $S_B^{(>2)}$ do not play a role in this scenario, and therefore we will omit this term altogether in the remainder of this chapter. Nevertheless, one ought to keep in mind that a treatment of particle-exciton or exciton-exciton interactions would require the inclusion of these kinds of diagrams.

With the above remark in mind, we now proceed to analyze the terms $(\mathcal{R} | [V^{qc} - \mathcal{U}]^{-1} | \mathcal{R})$ and $\operatorname{Tr} \log[-(V^{qc} - \mathcal{U})]$. By definition, $\mathcal{R}$ and $\mathcal{U}$ must be identified by looking at the diagrams in Eq. (4.24) that are linear and quadratic in $\delta\rho$, respectively. Let $\mathcal{R} = \sum_{n=1}^{\infty} \mathcal{R}^{(n)}$ and $\mathcal{U} = \sum_{n=0}^{\infty} \mathcal{U}^{(n)}$, where $\mathcal{R}^{(n)}$ and $\mathcal{U}^{(n)}$ contain all diagrams with exactly n instances of $\phi_B^* \phi_B$. Note that $\mathcal{R}^{(0)} = 0$, as there are no diagrams that are linear in $\delta\rho$ and do not contain bulk fields. Since in this chapter we only want to keep terms of order at most $(\phi_B^* \phi_B)^2$, we see that $(\mathcal{R} | [V^{qc} - \mathcal{U}]^{-1} | \mathcal{R})$ can simply be replaced by $(\mathcal{R}^{(1)} | [V^{qc} - \mathcal{U}^{(0)}]^{-1} | \mathcal{R}^{(1)})$. On the other hand, for the term in Eq. (4.31b) it suffices to take $\mathcal{U} = \sum_{n=0}^{2} \mathcal{U}^{(n)}$. The expressions for $\mathcal{R}^{(1)}$ and $\mathcal{U}^{(n)}$ for $n \leq 2$ are given in App. 4.A.

Using Eq. (4.31a) and the Feynman rules given in the appendix, one can show that the term $(\mathcal{R}^{(1)} | [V^{qc} - \mathcal{U}^{(0)}]^{-1} | \mathcal{R}^{(1)})$ is the following sum of bubble diagrams:

$$\frac{1}{2}(\mathcal{R}^{(1)} | [V^{qc} - \mathcal{U}^{(0)}]^{-1} | \mathcal{R}^{(1)}) \tag{4.32}$$

$$= \int \left\{ \vcenter{\hbox{}} + \vcenter{\hbox{}} + \cdots \right\}.$$

Together with $\frac{1}{2}\left(\phi_B^*\phi_B\big|V+V^{(1)}\big|\phi_B^*\phi_B\right)$, which contains the zero- and single-bubble diagrams, these terms contain two-body diagrams consisting of a chain of simple bubbles. We collectively denote them by $\frac{1}{2}\left(\phi_B^*\phi_B\big|V^{(\text{RPA})}\big|\phi_B^*\phi_B\right)$. The subscript "RPA" stands for the so-called *random-phase approximation*, which is detailed in the next section.

On the other hand, the effect of the trace operation on the terms in Eq. (4.31b) is to produce all diagrams that can be formed with the following pieces by joining together pairs of dots:

$$\tag{4.33}$$

Besides some unimportant vacuum bubble diagrams, this produces some self-energy and vertex corrections to the propagator, such as:

$$, \text{ etc.} \tag{4.34}$$

It also gives self-energy and vertex corrections to the two-body interaction, such as:

$$, \text{ etc.} \tag{4.35}$$

Note that not all of these corrections are generated in this way, but only those that are obtained from diagrams with exactly two dots. This is of course due to the fact that we have neglected contributions of order $\delta\rho^3$ and higher. Also, if we were interested in more-body interactions, we would have to include in the list of Eq. (4.33) similar diagrams with a higher amount of external bulk fields, which we have neglected for the purposes of this chapter.

In this way, we finally arrive at the following effective action, which is suitable for the analysis of the two-body interactions in Bi$_2$Se$_3$ nanosheets:

$$S_{B,\text{eff}}^{(\leq 2)}[\phi_B^*,\phi_B] = \left(\phi_B\big|-G_B[\rho_0]^{-1}\big|\phi_B\right) + \frac{1}{2}\left(\phi_B^*\phi_B\big|V^{(\text{RPA})}\big|\phi_B^*\phi_B\right)$$
$$+ \int\left\{\text{self-energy and vertex corrections arising from (4.33)}\right\}. \tag{4.36}$$

In order to further simplify the treatment, we now perform one of the most widely used approximations in the literature of many-body physics.

4.3 The random-phase approximation

In the so-called *random-phase approximation* (RPA), the self-energy and vertex corrections are neglected and the total effective interaction is approximated by the sum of all simple-bubble-chain diagrams such as those in Eq. (4.32) [98, Ch. 2, 115, Ch. 10]. In other words, we simply approximate the effective two-body action by

$$S_{B,\text{eff}}^{(\leq 2)}[\phi_B^*,\phi_B] \approx \left(\phi_B\big|-G_B[\rho_0]^{-1}\big|\phi_B\right) + \frac{1}{2}\left(\phi_B^*\phi_B\big|V^{(\text{RPA})}\big|\phi_B^*\phi_B\right). \tag{4.37}$$

The RPA is useful because the resulting diagrams form a geometric series which can be summed exactly. Indeed, if we denote the RPA interaction by a thick photon line, it satisfies the following diagrammatic Dyson equation:

$$\underset{\boldsymbol{x},z,\tau \qquad \boldsymbol{x}',z',\tau'}{} = \underset{\boldsymbol{x},z,\tau \qquad \boldsymbol{x}',z',\tau'}{} + \underset{\boldsymbol{x},z,\tau \qquad \boldsymbol{x}',z',\tau'}{} . \tag{4.38}$$

(Note that the indices $\mathbf{i} = \mathbf{i}'$ have been omitted everywhere.) In two-dimensional momentum and frequency space, the Dyson equation analytically reads

$$V^{(\mathrm{RPA})}(\boldsymbol{q}, \mathrm{i}\omega_n; z, z') = V(\boldsymbol{q}; z, z')$$
$$+ \int \mathrm{d}z_1 \int \mathrm{d}z_2\, V^{(\mathrm{RPA})}(\boldsymbol{q}, \mathrm{i}\omega_n; z, z_1) \left\{ \sum_{s,s'} \pi_s(z_1)\, \Pi_{ss'}(\boldsymbol{q}, \mathrm{i}\omega_n)\, \pi_{s'}(z_2) \right\} V(\boldsymbol{q}; z_2, z'). \tag{4.39}$$

The important object in this equation is the *polarization operator* $\Pi(\boldsymbol{q}, \mathrm{i}\omega_n)$, whose components are given by

$$\Pi_{ss'}(\boldsymbol{q}, \mathrm{i}\omega_n) = \frac{1}{V\beta} \sum_{\alpha,\alpha'} \sum_{\boldsymbol{k},m} G_{\mathrm{S}}[\rho_0]_{s,\alpha;s',\alpha'}(\boldsymbol{k}+\boldsymbol{q}, \mathrm{i}\omega_n + \mathrm{i}\omega_m)\, G_{\mathrm{S}}[\rho_0]_{s',\alpha';s,\alpha}(\boldsymbol{k}, \mathrm{i}\omega_m). \tag{4.40}$$

Note that $G_{\mathrm{S}}[\rho_0]$ contains the surface self-energy correction $\Sigma_{\mathrm{S}}[\rho_0]$ given in Eq. (4.19). We will only focus on homogeneous systems with constant average surface densities, $\langle \rho_s(\boldsymbol{x}, \tau) \rangle \equiv \rho_{0,s}$, in which case we obtain

$$\Sigma_{\mathrm{S}}[\rho_0]_{\mathbf{ss}'}(\boldsymbol{x}, \tau; \boldsymbol{x}', \tau') = \delta_{\mathbf{ss}'}\, \delta(\boldsymbol{x}-\boldsymbol{x}')\, \delta(\tau-\tau') \sum_{s''} \rho_{0,s''} \int \mathrm{d}^2 x''\, V_{ss''}^{\mathrm{qc}}(\boldsymbol{x}-\boldsymbol{x}''). \tag{4.41}$$

This Hartree self-energy can be canceled against a positively charged ion background, as in the well-known jellium model [90, Ch. 8]. For this reason we will neglect it henceforth.

4.3.1 The polarization operator

Inverting (4.14) in momentum space allows for the explicit calculation of the polarization. We will allow for different chemical potentials on both surfaces by changing $\mu \to \mu_s$, thus taking into account the possibility of gating the nanosheet. Note that, since the top and bottom surfaces are uncoupled from each other in the effective surface Hamiltonian, the polarization matrix has the diagonal structure $\mathrm{diag}(\Pi_t, \Pi_b)$. Thus, in principle one could easily account for potential tunneling effects by including a small coupling between the top and bottom surfaces. For simplicity, we neglect these effects in the present treatment. The only nonzero components of the polarization operator then read

$$\Pi_s(\boldsymbol{q}, \mathrm{i}\omega_n) = \frac{2}{V\beta} \sum_{\boldsymbol{k},m} \frac{\mathcal{A}_s(\boldsymbol{k}+\boldsymbol{q}, \mathrm{i}\omega_n + \mathrm{i}\omega_m)\mathcal{A}_s(\boldsymbol{k}, \mathrm{i}\omega_m) + \mathrm{Re}[\mathcal{B}(\boldsymbol{k}+\boldsymbol{q})\mathcal{B}^*(\boldsymbol{k})]}{\left[\mathcal{A}_s(\boldsymbol{k}+\boldsymbol{q}, \mathrm{i}\omega_n + \mathrm{i}\omega_m)^2 - |\mathcal{B}(\boldsymbol{k}+\boldsymbol{q})|^2\right] \left[\mathcal{A}_s(\boldsymbol{k}, \mathrm{i}\omega_m)^2 - |\mathcal{B}(\boldsymbol{k})|^2\right]}, \tag{4.42}$$

where

$$\mathcal{A}_s(\boldsymbol{k}, \mathrm{i}\omega_n) = -\mathrm{i}\omega_n + E_0 + D|\boldsymbol{k}|^2 - \mu_s, \tag{4.43a}$$
$$\mathcal{B}(\boldsymbol{k}) = -\mathrm{i}v_{\mathrm{F}}k_+. \tag{4.43b}$$

After performing a partial fraction decomposition of the denominator, the Matsubara sum can be evaluated by standard methods (cf. Ref. [90, Ch. 7]) and gives

$$\Pi_s(\boldsymbol{q}, \mathrm{i}\omega_n) = \tag{4.44}$$

$$\frac{1}{2V}\sum_{\boldsymbol{k}}\left[1+\frac{(\boldsymbol{k}+\boldsymbol{q})\cdot\hat{\boldsymbol{k}}}{\sqrt{k^2+q^2+2\boldsymbol{k}\cdot\boldsymbol{q}}}\right]\left[\frac{N_+^s(\boldsymbol{k}+\boldsymbol{q})-N_+^s(\boldsymbol{k})}{-\mathrm{i}\omega_n+\epsilon_+(\boldsymbol{k}+\boldsymbol{q})-\epsilon_+(\boldsymbol{k})}+\frac{N_-^s(\boldsymbol{k}+\boldsymbol{q})-N_-^s(\boldsymbol{k})}{-\mathrm{i}\omega_n+\epsilon_-(\boldsymbol{k}+\boldsymbol{q})-\epsilon_-(\boldsymbol{k})}\right]$$

$$+\frac{1}{2V}\sum_{\boldsymbol{k}}\left[1-\frac{(\boldsymbol{k}+\boldsymbol{q})\cdot\hat{\boldsymbol{k}}}{\sqrt{k^2+q^2+2\boldsymbol{k}\cdot\boldsymbol{q}}}\right]\left[\frac{N_+^s(\boldsymbol{k}+\boldsymbol{q})-N_-^s(\boldsymbol{k})}{-\mathrm{i}\omega_n+\epsilon_+(\boldsymbol{k}+\boldsymbol{q})-\epsilon_-(\boldsymbol{k})}+\frac{N_-^s(\boldsymbol{k}+\boldsymbol{q})-N_+^s(\boldsymbol{k})}{-\mathrm{i}\omega_n+\epsilon_-(\boldsymbol{k}+\boldsymbol{q})-\epsilon_+(\boldsymbol{k})}\right],$$

where $\hat{\boldsymbol{k}}$ is a unit vector in the direction of $\boldsymbol{k}$, $\epsilon_\pm(\boldsymbol{k})=\pm v_{\mathrm{F}}|\boldsymbol{k}|$ are the eigenenergies of the single-particle surface Hamiltonian, and $N_\pm^s(\boldsymbol{k})\equiv N_{\mathrm{FD}}(\epsilon_\pm(\boldsymbol{k})-\mu_s)$ with $N_{\mathrm{FD}}(x)=(e^{\beta x}+1)^{-1}$ are the Fermi-Dirac distributions of the conduction and valence bands, respectively. Note that we have neglected the term proportional to k^2 in the effective surface Hamiltonian, which only measures small deviations from the linear dispersion relation and is hence irrelevant in the low-energy limit. We have also set the origin of energy at the Dirac point, which simply amounts to a shift by E_0. The above equation can be written in the compact form

$$\Pi_s(\boldsymbol{q},\mathrm{i}\omega_n)=\frac{1}{V}\sum_{\boldsymbol{k},\kappa,\kappa'}\frac{N_\kappa^s(\boldsymbol{k}+\boldsymbol{q})-N_{\kappa'}^s(\boldsymbol{k})}{-\mathrm{i}\omega_n+\epsilon_\kappa(\boldsymbol{k}+\boldsymbol{q})-\epsilon_{\kappa'}(\boldsymbol{k})}F_{\kappa\kappa'}(\boldsymbol{k},\boldsymbol{k}+\boldsymbol{q}),\tag{4.45}$$

where $\kappa,\kappa'\in\{+,-\}$ and $F_{\kappa\kappa'}(\boldsymbol{k},\boldsymbol{k}')=\frac{1}{2}(1+\kappa\kappa'\cos\theta)$ is the overlap of states, with θ the angle between $\boldsymbol{k}$ and $\boldsymbol{k}'$. In this expression, the terms with $\kappa=\kappa'$ represent the system's response to intraband (valence-valence or conduction-conduction) transitions, whereas the terms with $\kappa\neq\kappa'$ correspond to interband (valence-conduction) transitions.

Eq. (4.45) has been evaluated explicitly in the zero-temperature limit in Refs. [116–118]. Note that the polarization in Ref. [118] is defined with the opposite sign, and furthermore in both papers it differs from (4.45) by a factor of $g_s g_v=4$ which accounts for the spin and valley degeneracy in graphene. In our case, the surface degeneracy $g_{\mathrm{S}}=2$ does not appear in the above equation because we treat both surfaces separately. To translate the results of the articles to our problem, we must change $g_s g_v$ to g_{S} and (in the case of [118]) γ to v_{F}, resulting in the density of states (DOS)

$$D(\varepsilon)=\sum_s D_s(\varepsilon)=\frac{g_{\mathrm{S}}}{2\pi}\frac{|\varepsilon|}{v_{\mathrm{F}}^2},\tag{4.46}$$

with $D_s(\varepsilon)=D(\varepsilon)/g_{\mathrm{S}}$ the DOS of a single surface.

For $\mu_s=0$ one finds a nonzero contribution due to the interband transitions, since even when the chemical potential is located at the Dirac point the gapless surface-state spectrum allows transitions between the valence and conduction bands. In this case the full polarization assumes the form [117]

$$\Pi_s(\boldsymbol{q},\omega)\Big|_{\mu_s=0}=-\frac{q^2}{16\sqrt{v_{\mathrm{F}}^2 q^2-\omega^2}}\Theta(v_{\mathrm{F}}q-\omega)-\mathrm{i}\frac{q^2}{16\sqrt{\omega^2-v_{\mathrm{F}}^2 q^2}}\Theta(\omega-v_{\mathrm{F}}q).\tag{4.47}$$

When $\mu_s\neq 0$ we define $\widetilde{\Pi}_s(\boldsymbol{q},\omega)\equiv\Pi_s(\boldsymbol{q},\omega)/D_s(\mu_s)$, where the chemical potential μ_s is just the Fermi energy of the surface s in the zero-temperature limit. We also introduce the dimensionless quantities $x\equiv q/k_{\mathrm{F},s}$ and $\nu\equiv\omega/\mu_s$ for each surface (we omit the surface subscript in x and ν to avoid clutter). Here, $k_{\mathrm{F},s}$ is the Fermi momentum of the surface s, defined by $\mu_s=\epsilon(k_{\mathrm{F},s})$. With these conventions, our $\widetilde{\Pi}_s$ is the same as $\widetilde{\Pi}$ in Ref. [118] except for the minus sign. In order to give an expression for the polarization in the general case $\mu_s\neq 0$ we define the complex function

$$G(z)=z\sqrt{z^2-1}-\log\left(z+\sqrt{z^2-1}\right).\tag{4.48}$$

Then, $\widetilde{\Pi}_s$ can be written as [117]

$$\widetilde{\Pi}_s(x,\nu)=-1+\frac{x^2}{8\sqrt{\nu^2-x^2}}\left\{\left[G\left(\frac{\nu+2}{x}\right)-\mathrm{i}\pi\right]-\Theta\left(\frac{2-\nu}{x}-1\right)\left[G\left(\frac{2-\nu}{x}\right)-\mathrm{i}\pi\right]\right.$$

$$-\Theta\left(\frac{\nu-2}{x}+1\right)G\left(\frac{\nu-2}{x}\right)\Bigg\}, \qquad (4.49)$$

where Θ is the Heaviside theta function. We can also give separate expressions for the real and imaginary parts of $\widetilde{\Pi}_s$ as done in [118]. The polarization is split as

$$\widetilde{\Pi}_s(x,\nu) = \widetilde{\Pi}_s^+(x,\nu) + \widetilde{\Pi}_s^-(x,\nu), \qquad (4.50)$$

where $\widetilde{\Pi}_s^+$ and $\widetilde{\Pi}_s^-$ correspond to the intraband and interband transitions, respectively. Furthermore, we write

$$\widetilde{\Pi}_s^+(x,\nu) = \widetilde{\Pi}_{s,1}^+(x,\nu)\,\Theta(\nu-x) + \widetilde{\Pi}_{s,2}^+(x,\nu)\,\Theta(x-\nu). \qquad (4.51)$$

The real parts of $\widetilde{\Pi}_{s,i}^+$ are given by

$$\text{Re}\left[\widetilde{\Pi}_{s,1}^+(x,\nu)\right] = -1 + \frac{1}{8\sqrt{\nu^2-x^2}}\Big\{f_1(x,\nu)\,\Theta(|\nu+2|-x)$$
$$+ \text{sgn}(\nu-2+x)\,f_1(x,-\nu)\,\Theta(|\nu-2|-x)$$
$$+ f_2(x,\nu)\left[\Theta(2+x-\nu)+\Theta(2-x-\nu)\right]\Big\}, \qquad (4.52\text{a})$$

$$\text{Re}\left[\widetilde{\Pi}_{s,2}^+(x,\nu)\right] = -1 + \frac{1}{8\sqrt{x^2-\nu^2}}\Big\{f_3(x,\nu)\,\Theta(x-|\nu+2|)$$
$$+ f_3(x,-\nu)\,\Theta(x-|\nu-2|)$$
$$+ \frac{\pi x^2}{2}\left[\Theta(|\nu+2|-x)+\Theta(|\nu-2|-x)\right]\Big\}, \qquad (4.52\text{b})$$

and their imaginary parts are given by

$$\text{Im}\left[\widetilde{\Pi}_{s,1}^+(x,\nu)\right] = \frac{1}{8\sqrt{\nu^2-x^2}}\Big\{f_3(x,-\nu)\,\Theta(x-|\nu-2|)$$
$$+ \frac{\pi x^2}{2}\left[\Theta(2+x-\nu)+\Theta(2-x-\nu)\right]\Big\}, \qquad (4.53\text{a})$$

$$\text{Im}\left[\widetilde{\Pi}_{s,2}^+(x,\nu)\right] = \frac{-1}{8\sqrt{x^2-\nu^2}}\Theta(2-x+\nu)\left[f_4(x,\nu)-f_4(x,-\nu)\,\Theta(2-x-\nu)\right]. \qquad (4.53\text{b})$$

On the other hand,

$$\widetilde{\Pi}_s^-(x,\nu) = -\frac{\pi x^2}{8\sqrt{x^2-\nu^2}}\Theta(x-\nu) - \mathrm{i}\frac{\pi x^2}{8\sqrt{\nu^2-x^2}}\Theta(\nu-x), \qquad (4.54)$$

which is just the dimensionless version of (4.47).

The functions f_i read as follows:

$$f_1(x,\nu) = (\nu+2)\sqrt{(\nu+2)^2-x^2} - x^2\log\left[\frac{(\nu+2)+\sqrt{(\nu+2)^2-x^2}}{|\nu+\sqrt{\nu^2-x^2}|}\right], \qquad (4.55\text{a})$$

$$f_2(x,\nu) = x^2\log\left(\frac{\nu-\sqrt{\nu^2-x^2}}{x}\right), \qquad (4.55\text{b})$$

$$f_3(x,\nu) = (\nu+2)\sqrt{x^2-(\nu+2)^2} + x^2\arcsin\left(\frac{\nu+2}{x}\right), \qquad (4.55\text{c})$$

$$f_4(x,\nu) = (\nu+2)\sqrt{(\nu+2)^2-x^2} - x^2\log\left[\frac{(\nu+2)+\sqrt{(\nu+2)^2-x^2}}{x}\right]. \qquad (4.55\text{d})$$

Together, the expressions presented in this section define the full analytic structure of the polarization operator in the zero-temperature limit.

4.3.2 Solving the Dyson equation

The Dyson equation giving the RPA potential, Eq. (4.39), cannot be solved analytically for an arbitrary interaction potential. However, we may exploit the fact that π_s are the probability densities of strongly localized surface states to solve it approximately to a very good degree of accuracy. Indeed, we may approximate these by

$$\pi_s(z) \approx \delta(z - \langle z \rangle_s), \tag{4.56}$$

where $\langle z \rangle_s = \int \mathrm{d}z\, z\, \pi_s(z)$ stands for the expected value of the z-coordinate. In practice, it suffices to take $\langle z \rangle_{t,b} \approx \pm \ell/2$ if the film is not unreasonably thin. Then the Dyson equation becomes an algebraic equation, namely

$$V^{(\mathrm{RPA})}(z, z') = V(z, z') + V^{(\mathrm{RPA})}_{\mathrm{B},t}(z)\, \Pi_t V_{\mathrm{B},t}(z') + V^{(\mathrm{RPA})}_{\mathrm{B},b}(z)\, \Pi_b V_{\mathrm{B},b}(z'). \tag{4.57}$$

For clarity we have suppressed the momentum and frequency dependencies, and introduced the notation $V_{\mathrm{B},s}(z) \equiv V(z, \langle z \rangle_s)$ meant to highlight that this in general represents the interaction between the surface s and a point in the bulk with coordinate z. This equation can be solved by first putting $z' = \langle z \rangle_s$ with $s \in \{t, b\}$ and solving a system of equations for $V^{(\mathrm{RPA})}_{\mathrm{B},s}(z)$ in terms of $V_{ss'} \equiv V(\langle z \rangle_s, \langle z \rangle_{s'})$. This intermediate result is

$$\begin{bmatrix} V^{(\mathrm{RPA})}_{\mathrm{B},t}(z) \\ V^{(\mathrm{RPA})}_{\mathrm{B},b}(z) \end{bmatrix} = \frac{1}{(1 - \Pi_t V_{tt})(1 - \Pi_b V_{bb}) - \Pi_t \Pi_b V_{tb}^2} \begin{bmatrix} 1 - \Pi_b V_{bb} & \Pi_b V_{tb} \\ \Pi_t V_{tb} & 1 - \Pi_t V_{tt} \end{bmatrix} \begin{bmatrix} V_{\mathrm{B},t}(z) \\ V_{\mathrm{B},b}(z) \end{bmatrix}. \tag{4.58}$$

Plugging these results back into the right-hand side of Eq. (4.57) directly gives the RPA-corrected bulk-bulk interaction $V^{(\mathrm{RPA})}(z, z')$.

We shall mostly concern ourselves with the case when the system enjoys inversion symmetry along the z-direction, meaning that $\Pi_t = \Pi_b \equiv \Pi$ as well as $V_{tt} = V_{bb}$. Then,

$$V^{(\mathrm{RPA})}_{\mathrm{B},t}(z) = V^{(\mathrm{RPA})}_{\mathrm{B},b}(-z) = \frac{(1 - \Pi V_{tt}) V_{\mathrm{B},t}(z) + \Pi V_{tb} V_{\mathrm{B},t}(-z)}{[1 - \Pi(V_{tt} - V_{tb})][1 - \Pi(V_{tt} + V_{tb})]}, \tag{4.59}$$

and the screened bulk-bulk potential becomes

$$V^{(\mathrm{RPA})}(z, z') = V(z, z') + \frac{\mathcal{P}(z, z') + \mathcal{P}(-z, -z')}{[1 - \Pi(V_{tt} - V_{tb})][1 - \Pi(V_{tt} + V_{tb})]}, \tag{4.60}$$

where

$$\mathcal{P}(z, z') \equiv \Pi(1 - \Pi V_{tt}) V_{\mathrm{B},t}(z) V_{\mathrm{B},t}(z') + \Pi^2 V_{tb} V_{\mathrm{B},t}(z) V_{\mathrm{B},t}(-z'). \tag{4.61}$$

We have now determined both the polarization and the effective RPA interaction. Therefore, we are ready to investigate different kinds of many-body effects in bismuth selenide.

4.4 Static screening

In this section we wish to analyze how the topological surface states modify the bulk interaction potential between particles in a thin slab. In general the polarization function depends on the frequency of the particles involved in the scattering and hence introduces retardation effects to the effective screened interaction potential. However, in the low-energy limit we expect the typical times between successive scattering events to be much longer than the characteristic

time scale associated with a photon exchange. Hence we are mainly interested in the zero-frequency limit of the effective interaction, also known as the *static limit*. In this case, the components of the polarization function at zero temperature are [118]:

$$\widetilde{\Pi}_s^+(x,0) = \begin{cases} -1 + \dfrac{\pi x}{8}, & x \leq 2, \\ -1 + \dfrac{1}{2}\sqrt{1 - \dfrac{4}{x^2}} + \dfrac{x}{4}\arcsin\left(\dfrac{2}{x}\right), & x > 2, \end{cases} \tag{4.62a}$$

$$\widetilde{\Pi}_s^-(x,0) = -\dfrac{\pi x}{8}. \tag{4.62b}$$

We see that the total static polarization of a single surface becomes a constant at $q \leq 2k_F$, just like in a normal 2D system with parabolic band dispersions [119]. More precisely, in this case Π_s is (minus) the density of states at the Fermi level of the surface s, namely

$$\Pi_s(q,0)\Big|_{q \to 0} = -\dfrac{1}{2\pi}\dfrac{|\mu_s|}{v_F^2}. \tag{4.63}$$

From the point of view of the exciton problem considered in Ch. 4, the static approximation is relevant because the typical frequency exchange in the direct interaction channel is expected to be small. Therefore, the bare interaction potential in this channel ought to be modified by its statically screened counterpart in order to obtain more accurate results. We now proceed to analyze the static screening of the potentials of interest in the event of a small but nonzero chemical potential. We will work in the long-wavelength limit in order to use the above constant expression as the total polarization at all times. Note that this is in principle not correct when finding the screened real-space potentials by performing a Fourier transform, which is a sum over *all* values of q. Nevertheless, this so-called *Thomas-Fermi screening* is widely discussed in the condensed-matter literature and counts with great experimental success [120–122]. For simplicity we will set $\mu_t = \mu_b$ in the remainder of this section.

4.4.1 Homogeneous interaction potentials

We begin with the simpler case when the bare interaction between particles anywhere within the nanosheet is the same independently of their z coordinates, that is, $V_{tt} = V_{tb} = V_{B,t} = V_{B,b} \equiv V$. This is a good approximation when the film thickness is very small and the typical momenta satisfy $q\ell \ll 1$. In this case Eq. (4.60) reduces to

$$V^{(\mathrm{RPA})}(q,\omega) = \dfrac{V(q)}{1 - 2\Pi(q,\omega)V(q)}, \tag{4.64}$$

which is the standard RPA result for the case of an interaction V, with a factor of 2 arising from the fact that there are two surfaces [98, Ch. 2, 115].

For convenience, we define a characteristic system screening length r_0 as well as a characteristic bulk energy scale V_0 through

$$r_0 \equiv \dfrac{\varepsilon_d}{2\varepsilon_s}\ell, \tag{4.65a}$$

$$V_0 \equiv \dfrac{e^2}{4\pi\varepsilon_0\varepsilon_d\ell}. \tag{4.65b}$$

We will study the screening of both the Coulomb and the Struve-Neumann potentials. In the above units, they are expressed in real and momentum space as

$$V^C(r) = 2V_0\dfrac{r_0}{r}, \qquad\qquad V^C(q) = r_0^2 V_0 \dfrac{4\pi}{r_0 q}, \tag{4.66}$$

$$V^{\text{SN}}(r) = \pi V_0 \left[\mathbf{H}_0\left(\frac{r}{r_0}\right) - Y_0\left(\frac{r}{r_0}\right) \right], \qquad V^{\text{SN}}(q) = r_0^2 V_0 \frac{4\pi}{r_0 q (1 + r_0 q)}. \tag{4.67}$$

(The definition of r_0 in fact arises from the characteristic decay length of the Struve-Neumann potential.) We also introduce a screening length λ through

$$\lambda^{-1} \equiv \frac{1}{2\pi} \frac{e^2}{\varepsilon_0 \varepsilon_s} \frac{|\mu|}{v_{\text{F}}^2}. \tag{4.68}$$

We will see that the screened interactions of interest are controlled by the single dimensionless parameter

$$\xi \equiv \frac{r_0}{\lambda}. \tag{4.69}$$

First we consider the Coulomb potential. Incorporating the Thomas-Fermi screening gives

$$V^{\text{C(TF)}}(q) = r_0^2 V_0 \frac{4\pi}{r_0 q + \xi}. \tag{4.70}$$

The inverse Fourier transform can be calculated and the screened Coulomb potential in real space reads

$$V^{\text{C(TF)}}(r) = 2V_0 \frac{r_0}{r} \left\{ 1 - \xi \frac{\pi}{2} \frac{r}{r_0} \left[\mathbf{H}_0\left(\xi \frac{r}{r_0}\right) - Y_0\left(\xi \frac{r}{r_0}\right) \right] \right\}. \tag{4.71}$$

Again, $\mathbf{H}_0$ and Y_0 are the Struve and Neumann functions, respectively. At short distances one recovers the usual Coulomb potential. By contrast, when $\xi > 0$ the effects of screening become apparent at large distances. Using the asymptotic expansion

$$\left[\mathbf{H}_0(x) - Y_0(x) \right]\Big|_{x \to \infty} = \frac{2}{\pi x} - \frac{2}{\pi x^3} + \mathcal{O}\left(\frac{1}{x^5}\right), \tag{4.72}$$

one finds

$$V^{\text{C(TF)}}(r)\Big|_{r \to \infty} = \frac{2V_0}{\xi^2} \left(\frac{r_0}{r}\right)^3. \tag{4.73}$$

Within the Thomas-Fermi approximation to the RPA potential, the effects of screening cause the Coulomb potential to decay as r^{-3} at large distances, instead of the usual r^{-1}. However, the dependence in r is still algebraic in 2D, in stark constrast with the three-dimensional case where the screened Thomas-Fermi interaction decays exponentially as $\mathrm{e}^{-r/\lambda}/r$.

Next we study what happens to the physically more relevant Struve-Neumann potential. Its screened version in momentum space is given by

$$V^{\text{SN}}(q) = r_0^2 V_0 \frac{4\pi}{r_0 q (1 + r_0 q) + \xi}, \tag{4.74}$$

and in real space it is determined by the following integral:

$$V^{\text{SN(TF)}}(r) = 2V_0 \int_0^\infty \mathrm{d}x \, \frac{x J_0(x)}{x^2 + (r/r_0)x + \xi(r/r_0)^2}, \tag{4.75}$$

with $J_0(x)$ the Bessel function of the first kind. The integral can be performed by doing a partial fraction decomposition of the denominator, giving

$$V^{\text{SN(TF)}}(r) = \frac{\pi V_0}{\sqrt{1 - 4\xi}} \left\{ \Xi_+ \left[\mathbf{H}_0\left(\Xi_+ \frac{r}{r_0}\right) - Y_0\left(\Xi_+ \frac{r}{r_0}\right) \right] - \Xi_- \left[\mathbf{H}_0\left(\Xi_- \frac{r}{r_0}\right) - Y_0\left(\Xi_- \frac{r}{r_0}\right) \right] \right\}, \tag{4.76}$$

where $\Xi_\pm \equiv \frac{1}{2}\left(1 \pm \sqrt{1-4\xi}\right)$. One may verify that this potential is real even for $\xi > 1/4$. The case $\xi = 1/4$ has to be treated separately by either taking the limit of the above expression or repeating the integral. The result is

$$V^{\mathrm{SN(TF)}}(r)\Big|_{\xi=1/4} = \pi V_0 \left\{ \left[\mathbf{H}_0\left(\frac{r}{2r_0}\right) - Y_0\left(\frac{r}{2r_0}\right)\right] + \frac{r}{2r_0}\left[\mathbf{H}_{-1}\left(\frac{r}{2r_0}\right) + Y_1\left(\frac{r}{2r_0}\right)\right]\right\}. \quad (4.77)$$

One may again use the asymptotic expansion (4.72) to verify that $V^{\mathrm{SN(TF)}}(r \to \infty)$ is identical to (4.73). This comes as no surprise, since the unscreened counterpart of the Struve-Neumann potential reduces to the usual Coulomb potential in this limit.

Fig. 4.1 shows a comparison between the screened and unscreened versions of the Coulomb and Struve-Neumann potentials derived in this section.
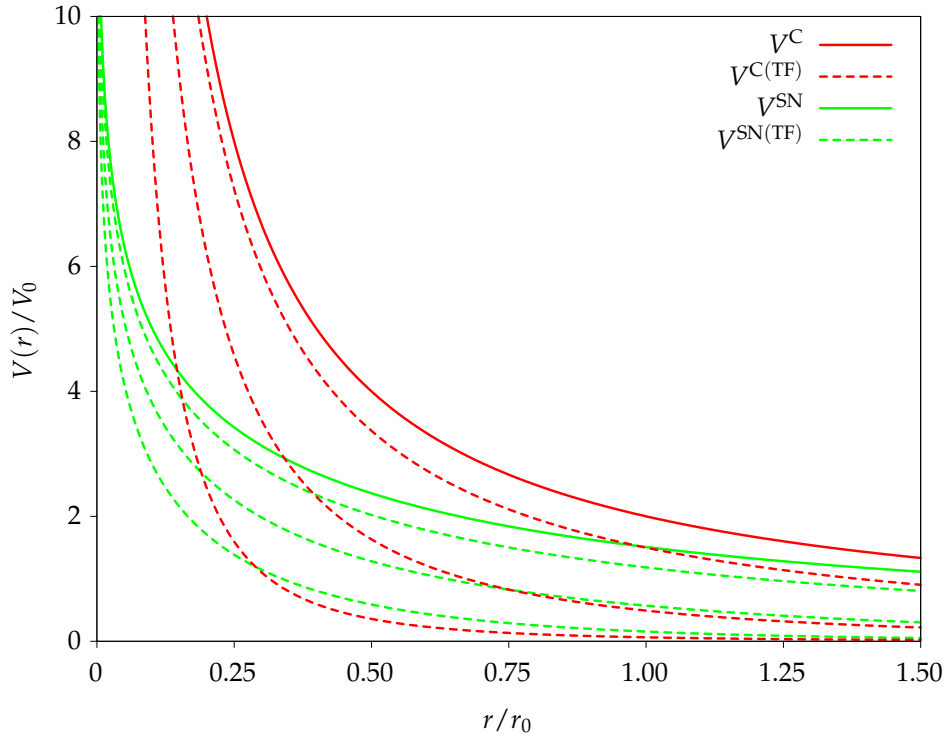
Figure 4.1. Comparison between the bare (solid lines) and Thomas-Fermi-screened (dashed lines) Coulomb and Struve-Neumann interaction potentials (red and green curves, respectively) for different values of ξ. In both cases, the dashed curves correspond (from top to bottom) to $\xi = 0.1, 1, 5$. The energies have been normalized to $V_0 = e^2/4\pi\varepsilon_0\varepsilon_d\ell$.

4.4.2 Rytova-Keldysh potential

Here we consider the static Thomas-Fermi screening of the full Rytova-Keldysh potential, which in momentum space reads

$$V^{\mathrm{RK}}(q; z, z') = 4\pi V_0 \ell^2 \frac{\cosh\left[q(\ell/2 - z) + \eta\right]\cosh\left[q(\ell/2 + z') + \eta\right]}{q\ell\sinh\left[q\ell + 2\eta\right]}, \quad (4.78)$$

In this case there are not two, but three competing length scales that determine the behavior of the screened interaction, namely ℓ, r_0, and λ. This is in contrast to the previous section, as the Coulomb and Struve-Neumann potentials implicitly assume the limit $\ell \ll r_0$. Hence, we need two dimensionless parameters instead of a single one to fully describe the screening of the RK

potential. It will be convenient in this case to consider a slightly different screening length from that of the previous section, namely

$$\Lambda^{-1} = \frac{1}{2\pi} \frac{e^2}{\varepsilon_0 \varepsilon_d} \frac{|\mu|}{v_F^2} = \frac{\varepsilon_s}{\varepsilon_d} \lambda^{-1}, \tag{4.79}$$

and use the dimensionless parameter

$$\zeta \equiv \frac{\ell}{\Lambda}. \tag{4.80}$$

The second sensible parameter to choose is the ratio between the dielectric constants of the bulk and the environment, $\varepsilon_d / \varepsilon_s$.

Due to the z, z'-dependence of the RK potential, the screening must be determined from the full RPA expression of Eq. (4.60). We have done this numerically for different values of ζ, z, and z' with the dielectric constants $\varepsilon_d = 28$ and $\varepsilon_s = 6$ which were already used in the previous chapter. A comparison is shown in Fig. 4.2.
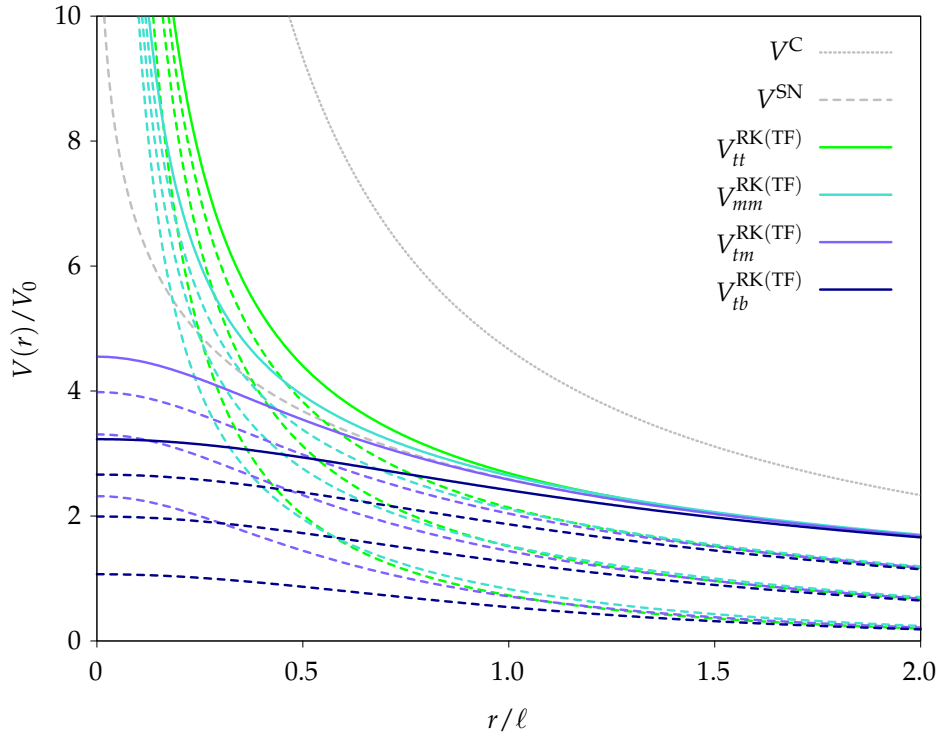
Figure 4.2. Comparison between the bare (solid lines) and Thomas-Fermi-screened (dashed lines) Rytova-Keldysh interaction potential for different values of ζ and different pairs (z, z'). The curves in the figure show the top-top, middle-middle, top-middle and top-bottom interactions within the nanosheet. In all cases, the dashed curves correspond (from top to bottom) to $\zeta = 0.02, 0.1, 0.5$. The energies have been normalized to $V_0 = e^2 / 4\pi \varepsilon_0 \varepsilon_d \ell$.

4.5 Plasmons in the RPA

The RPA calculation also allows us to identify collective quasiparticle excitations of the surface density fields. These collective excitations consist of fluctuations in the electron densities that propagate on the corresponding surface, and go by the name of *surface plasmons*. Looking at the last term in Eq. (4.28) we see that $V^{qc} - \mathcal{U}$ is the inverse surface density propagator. The quasiparticle excitations are therefore located at the poles of $(V^{qc} - \mathcal{U})^{-1}$, which according to

(4.32) essentially corresponds to $V^{(\mathrm{RPA})}$. We are thus interested in determining the poles of the *dielectric function*[1]

$$\epsilon(\boldsymbol{q}, \omega) \equiv [1 - \Pi(V_{tt} - V_{tb})][1 - \Pi(V_{tt} + V_{tb})]. \tag{4.81}$$

Since we are only interested in the long-wavelength behavior of the plasmon modes, it is justified to simply use the Coulomb interaction for points lying on the same surface, that is, $V_{tt}(q) = V^{\mathrm{C}}(q)$. The interaction between the top and bottom surfaces, however, is taken as $V_{tb}(q) = V^{\mathrm{C}}(q)\,e^{-q\ell}$. This result can be obtained by Fourier-transforming the three-dimensional Coulomb interaction after modifying the total distance between particles from r to $\sqrt{r^2 + \ell^2}$.

In the long-wavelength limit, $q \to 0$, the dimensionless polarization in the high- and low-frequency regimes assumes the following forms:

$$\widetilde{\Pi}(x, \nu) = \begin{cases} \dfrac{x^2 \nu^2}{2}\left(1 - \dfrac{\nu^2}{4}\right), & x < \nu < 2, \\[2ex] 1 + \mathrm{i}\dfrac{\nu}{x}, & \nu < x. \end{cases} \tag{4.82}$$

Solving the equation $1 - \Pi(V_{tt} + V_{tb}) = 0$ to first order in q yields the dispersion

$$\omega_+(q) = \omega_0 \sqrt{\dfrac{q}{k_{\mathrm{F}}}}, \quad \text{where} \quad \omega_0 \equiv \sqrt{\dfrac{g_{\mathrm{S}}}{8\pi}\dfrac{e^2 k_{\mathrm{F}}}{\varepsilon_0 \varepsilon_s}|\mu|}. \tag{4.83}$$

The corresponding plasmons are called *optical*, and correspond to the in-phase mode of the coupled two-surface system. On the other hand, the out-of-phase mode is realized by *acoustic* plasmons, whose dispersion is found from the equation $1 - \Pi(V_{tt} - V_{tb})$ and reads

$$\omega_-(q) = \omega_0 \sqrt{\dfrac{\ell k_{\mathrm{F}}}{2}}\dfrac{q}{k_{\mathrm{F}}}. \tag{4.84}$$

An illustration of both plasmon modes can be found in Fig. 4.3. The optical mode always exists in the long-wavelength limit, as $\omega_+(q)/q \to 0$ for $q \to 0$, and thus the condition $x < \nu$ which was used in Eq. (4.82) is always satisfied. However, the same cannot be said about the acoustic plasmons, because $\omega_-(q)$ is linear in the momentum and therefore we need to make sure that the proportionality constant is compatible with this precondition. The consequence of this is that, for a fixed value of ℓ, there exists a minimum chemical potential required to observe surface plasmons of this nature. One finds the condition $|\mu| > \mu^* \equiv \frac{8\pi v_{\mathrm{F}}}{\ell}\frac{e^2}{\varepsilon_0 \varepsilon_s}$, or equivalently, $\lambda < \ell/4$. For smaller $|\mu|$, the acoustic branch enters the region $\omega < v_{\mathrm{F}} q$ and becomes overdamped [123].

In Fig. 4.4 we show plots of the real and imaginary parts of $\epsilon(\boldsymbol{q}, \omega)$. Similarly, Fig. 4.5 shows plots of the real and imaginary parts of $\epsilon(\boldsymbol{q}, \omega)^{-1}$, which can be regarded as a spectral function. As such, its imaginary part is strongly peaked when $\mathrm{Re}[\epsilon(\boldsymbol{q}, \omega)] = 0$, that is, at the plasmon dispersions. This is clearly observed in the corresponding figure after broadening the dispersion by adding to the frequency a small imaginary part.

4.5.1 Plasmon decay

In the previous section we have determined the plasmon dispersions in the long-wavelength limit, where the polarization was a real function. However, generally speaking, Π will have an imaginary part and one must solve for $\epsilon(\boldsymbol{q}, \omega_p - \mathrm{i}\gamma) = 0$, where the usual real plasmon

---

[1]In most of the literature, the dielectric function is defined through $V^{(\mathrm{RPA})} = V/\epsilon$. It is clear from Eq. (4.60) that our definition of ϵ does not entirely conform to this custom. However, this is not a problem in view of the fact that our ϵ still carries the entirety of the singular behavior of $V^{(\mathrm{RPA})}$.
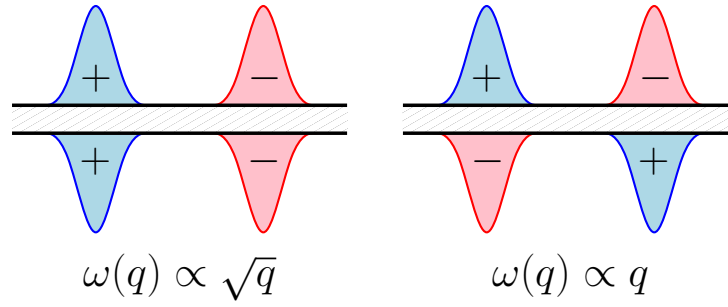
Figure 4.3. Cartoons of the in-phase (optical) and out-of-phase (acoustic) plasmons on the surfaces of a Bi_2Se_3 nanosheet. The bumps represent the charge-density oscillations, which must add up to zero on each surface due to electric neutrality.

dispersion ω_p acquires an additional imaginary piece γ [117]. The imaginary part of a dispersion relation is associated with a decay rate. This may be seen by observing that the time evolution of an eigenstate of complex energy $\varepsilon_n = \omega_n - i\gamma_n$ is given by $e^{-i\varepsilon_n t}|\psi_n\rangle$, and thus acquires a finite lifetime $\tau_n = (2\gamma_n)^{-1}$, as the probability of observation becomes proportional to e^{-t/τ_n}.

In our case, plasmons decay into surface electron-hole states. This decay is possible whenever the plasmon dispersion lies in the electron-hole continuum or single-particle excitation (SPE) region, defined by a nonzero imaginary part of the polarization function. Thus, in order to fully understand the decay mechanism, we must find the boundaries of said regions, that is, the allowed values of the energy change ω as a function of the momentum transfer q. Both intraband (valence-valence) and interband (valence-conduction) transitions between surface states are possible. Fig. 4.6 shows different scenarios and the precise boundaries are given in the caption.

In Fig. 4.7 we have again plotted the real and imaginary parts of the dielectric function, now on a larger scale, and we show the boundaries of the SPE regions explained above. We see by comparing the plots that the plasmon modes, defined by $\text{Re}[\epsilon(\boldsymbol{q},\omega)] = 0$, always enter one of the SPE regions. The optical mode always has a part inside the interband transition region and never in the intraband region, and thus this mode can only decay via the former. However, decay requires a minimum value of q, as the plasmon is not damped in the leftmost dark-blue region of Fig. 4.6 where the polarization is real. On the other hand, the acoustic mode will only decay via interband transitions if the chemical potential is higher than μ^*, but in the overdamped regime $|\mu| < \mu^*$ decay may take place via both intra- and interband transitions. Note that the figures all show the case $|\mu| > \mu^*$.

4.6 Corrections to the exciton spectra

The corrections to the bare interaction potential in the RPA scheme will in principle have implications on the exciton states we determined in Ch. 3. Due to the coupling with the surface electrons and holes, the correspoding exciton dispersions will in general acquire a complex energy shift. We can compute this shift by treating the quantity $\delta V \equiv V^{(\text{RPA})} - V$ as a perturbation, where

$$\delta V(\boldsymbol{q},\omega) \equiv -\frac{2\Pi(\boldsymbol{q},\omega)V(\boldsymbol{q})^2}{1 - 2\Pi(\boldsymbol{q},\omega)V(\boldsymbol{q})}. \tag{4.85}$$

Calculating the average value of this operator with the unperturbed wave functions then gives a first-order correction to the energy.

This complex energy shift will have two effects. Firstly, the real part will slightly modify the unperturbed energy levels. We must check that their shift preserves the original ordering of
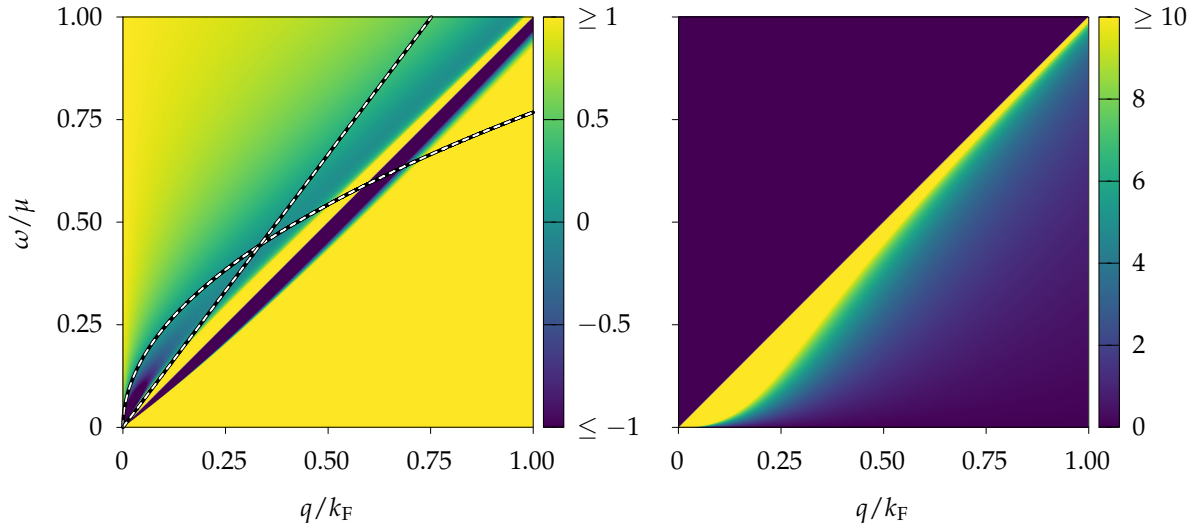
Figure 4.4. Left and right, respectively: typical plots of the real and imaginary parts of $\epsilon(\boldsymbol{q}, \omega)$. The dashed white lines indicate the local long-wavelength plasmon dispersions $\omega_+(q) \propto \sqrt{q}$ and $\omega_-(q) \propto q$ resulting from setting the real part equal to zero. We have chosen a sufficiently large value of μ such that the acoustic mode lies outside of the electron-hole continuum $\omega < v_F q$.
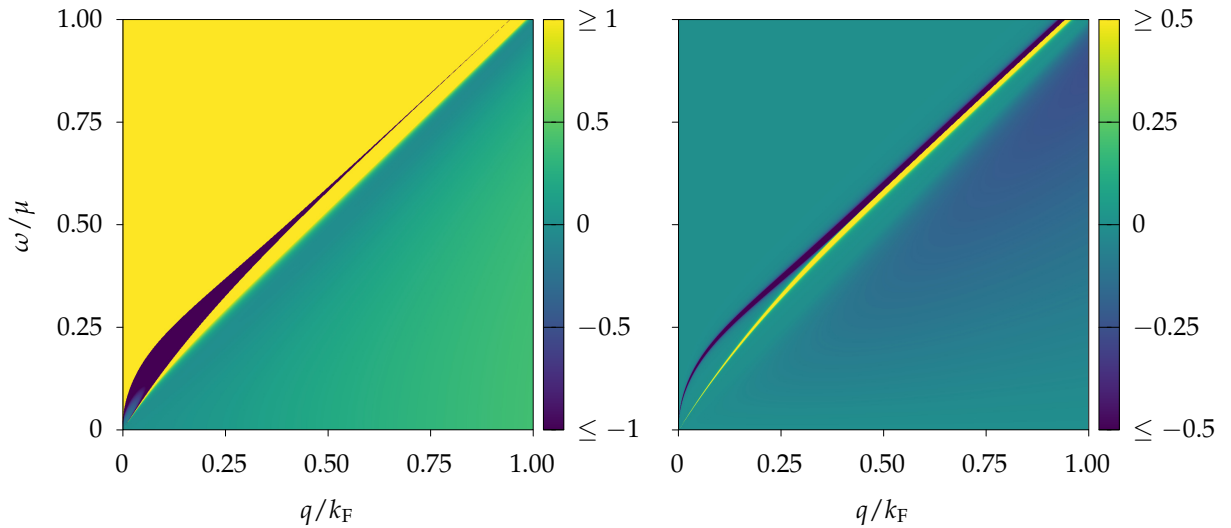
Figure 4.5. Left and right, respectively: plots of the real and imaginary parts of the spectral-like function $\epsilon(\boldsymbol{q}, \omega)^{-1}$. The plot of the imaginary part clearly shows the two plasmon dispersions. Formally these correspond to delta functions, but we have broadened them for visualization purposes by adding to ω a small imaginary part.

states and does not induce level crossings, which would correspond to nonperturbative effects. Secondly, the imaginary part will now give the excitons a finite lifetime. From a point of view of the dispersions, this corresponds to a broadening of the spectral lines. We must therefore check that the energy difference between successive levels is larger than their combined half-widths, since otherwise both levels will become hard to distinguish in a spectroscopy experiment. In this section we have restricted ourselves to the results obtained with the Struve-Neumann potential.

Excitons interact via the direct and exchange channels, which are now proportional to the
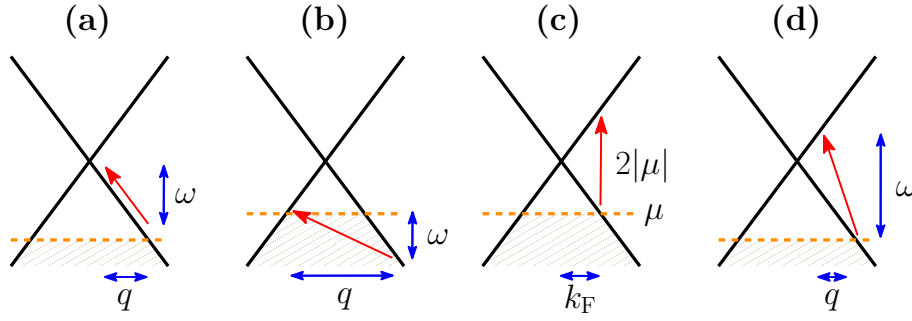
Figure 4.6. Different kinds of transitions between surface bands, represented by a red arrow. The chemical potential is represented as a dashed orange line and satisfies $|\mu| = v_F k_F$. **(a)** Intraband transitions with $q < 2k_F$ are only allowed in the range $0 < \omega < v_F q$. **(b)** Intraband transitions with $q > 2k_F$, on the other hand, are only allowed for $\omega > v_F(q - 2k_F)$. **(c)** Direct ($q = 0$) interband transitions are forbidden for $0 < \omega < 2|\mu|$. **(d)** In general for $q < 2k_F$, interband transitions are allowed only for $\omega > v_F(2k_F - q)$.
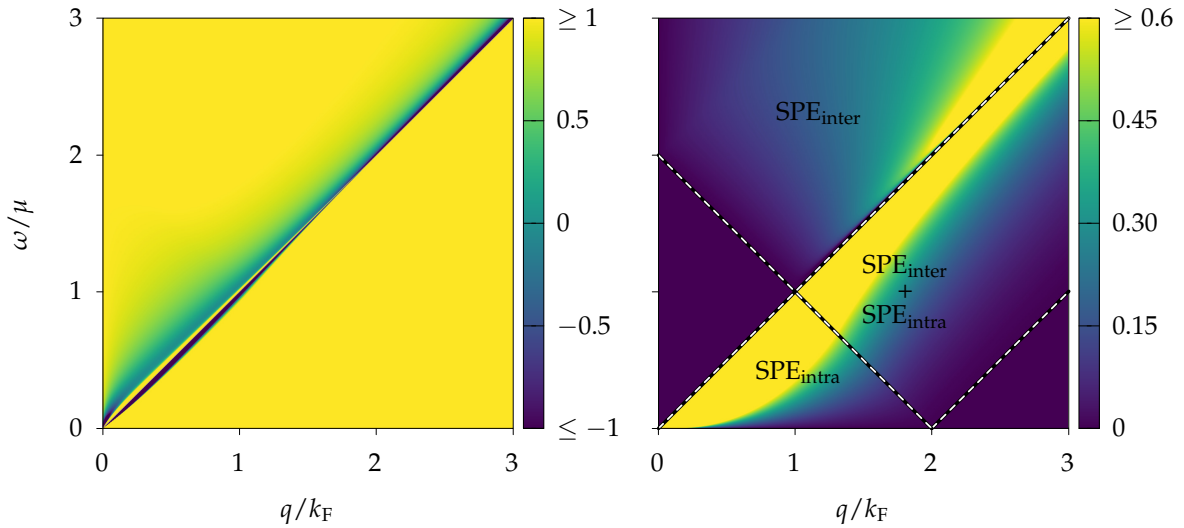
Figure 4.7. Left and right, respectively: typical plots of the real and imaginary parts of $\epsilon(q, \omega)$. The right panel shows the allowed transitions between surface bands. No transitions are possible inside the dark blue regions, where $\text{Im}[\epsilon(q, \omega)] = 0$. Plasmon decay can take place whenever the plasmon mode lies within one of the allowed transition regions.

interaction potentials $V^{(\text{RPA})}(k - k', \omega - \omega')$ and $V^{(\text{RPA})}(Q, \varepsilon_Q)$, respectively. Here, the internal frequencies ω and ω' are defined analogously to k and k', i.e., by defining the electron and hole energies as $\varepsilon_Q/2 + \omega$ and $\varepsilon_Q/2 - \omega$, respectively. We will employ the static approximation for the exchange diagram by replacing $\omega - \omega'$ by its mean value, which is simply zero. The direct potential is then real and does not contribute to a finite exciton lifetime but only to a real energy shift. Therefore, and because the results obtained in Fig. 3.2 indicate that $\varepsilon_Q > v_F Q$ for all Q, excitons can only decay through the exchange channel in this approximation. However, the exciton eigenstates in the subspaces $s \neq t$ do not perceive the exchange interaction, and thus their lifetimes remain infinite in the static approximation. Note that this result is exact, as the RPA only modifies the electrostatic interaction and does not introduce couplings between the different subspaces. The effect on the eigenstates living in the subspace with $s = t$ is most easily studied by considering Q along the x-direction and the gauge $\Phi_{Q\hat{x}, ++}(k) = \pm \Phi^*_{Q\hat{x}, --}(k)$, where the $+$ and $-$ signs correspond to the higher- and lower-energy eigenstates, respectively.

In this case, the expected value of the perturbation is given by

$$\langle \delta \hat{V}^X \rangle \Big|_{s=t} = -V^{(\text{RPA})}(\mathbf{Q}, \varepsilon_Q) \Big| \frac{1}{V} \sum_{\mathbf{k}} \Big\{ \langle \chi^{v,+}_{-Q\hat{x}/2+\mathbf{k}} | \chi^{c,+}_{Q\hat{x}/2+\mathbf{k}} \rangle \Phi_{Q\hat{x},++}(\mathbf{k})$$
$$\pm \langle \chi^{v,+}_{-Q\hat{x}/2+\mathbf{k}} | \chi^{c,+}_{Q\hat{x}/2+\mathbf{k}} \rangle^* \Phi^*_{Q\hat{x},++}(\mathbf{k}) \Big\} \Big|^2. \qquad (4.86)$$

The terms inside brackets reduce to the imaginary part of $\langle \chi^{v,+}_{-Q\hat{x}/2+\mathbf{k}} | \chi^{c,+}_{Q\hat{x}/2+\mathbf{k}} \rangle \Phi_{Q\hat{x},++}(\mathbf{k})$ for the lower state. However, in view of the property

$$\langle \chi^{v,+}_{-Q\hat{x}/2+\mathbf{k}_\phi} | \chi^{c,+}_{Q\hat{x}/2+\mathbf{k}_\phi} \rangle = \langle \chi^{v,+}_{-Q\hat{x}/2+\mathbf{k}_{-\phi}} | \chi^{c,+}_{Q\hat{x}/2+\mathbf{k}_{-\phi}} \rangle^*, \qquad (4.87)$$

the total integral must be real, as can be seen by again expanding $\Phi_{Q\hat{x},++}(\mathbf{k})$ into Fourier modes with real coefficients. This means that there is no shift for the lower exciton mode to first order in perturbation theory. By contrast, because the real part does not integrate to zero, the upper mode does pick up a nonzero first-order correction (unless $Q = 0$, in which case the exchange interaction vanishes). However, we have numerically verified that the corresponding real energy shifts are several orders of magnitude smaller than those arising from the direct diagram, and that the complex parts give extremely long lifetimes of hundreds of nanoseconds, making the broadening completely negligible.

These results indicate that the shift of the excitonic energy levels will be mostly due to the direct channel, and that exciton decay might actually be dominated by deviations from the static approximation in the entire momentum range. While the latter are outside the scope of the present work, we have numerically calculated the former in order to verify that the obtained values are *a posteriori* compatible with a perturbative treatment. The results for $\mathbf{Q} = \mathbf{0}$ are shown in Table 4.1. We see that, in all cases, the energy correction is quite small compared to the unperturbed binding energies. More importantly, however, the corrected binding energies of the individual states preserve the unperturbed level ordering. For these reasons, the effect of Eq. (4.85) can be regarded as a true perturbation.

m	0	-1	1	-2	2	-3	3
$\Delta_m^{(0)}$ (meV)	25.0	21.8	19.3	18.0	16.1	15.3	13.9
$\delta\varepsilon_m$ (meV)	5.0	4.7	4.4	4.2	4.0	3.9	3.6
$\Delta_m^{(1)}$ (meV)	20.0	17.1	14.9	13.8	12.1	11.4	10.3

Table 4.1. Energy shifts $\delta\varepsilon_m$ for the exciton eigenstates $|0; +, +; 0, m\rangle$ and $|0; -, -; 0, -m\rangle$ as found for the Struve-Neumann potential with $\varepsilon_d = 28$ and $\varepsilon_s = 6$ at zero chemical potential. The shifts are upward due to the weaker nature of the screened RPA potential in the attractive direct channel. We also give the unperturbed binding energies $\Delta_m^{(0)}$ as well as their first-order-corrected values, $\Delta_m^{(1)} = \Delta_m^{(0)} - \delta\varepsilon_m$. We see that the corrected binding energies of the individual states are such that the original level orderings remain unaltered.

It is interesting to note that the corrected values of the binding energies are now smaller than the splitting between the first and second bulk states of Table 2.2, *both* for the conduction and valence bands. This is in contrast to their unperturbed counterparts, for which the valence-band splitting was slightly smaller in the case of the ground state, as previously discussed in Sec. 3.6. This implies that the lowest-bulk-subband projection procedure we employed in Ch. 2 to obtain a low-energy model is completely justified for the description of excitons when many-body screening effects are included.

The results of this section validate our approach to excitons in Ch. 3. The qualitative aspects of the obtained dispersion relations remain robust, i.e., excitons remain indirect and the modes preserve their linear or quadratic behavior at $\mathbf{Q} = \mathbf{0}$. The broadening of the spectral lines

in the static approximation is either identically zero or absolutely negligible in comparison to the typical energy scales. Clearly, exciton lifetimes must be estimated better by considering deviations from the static approximation, which is left for future work. We note that we expect the lifetimes of the different eigenstates to increase with energy. This can be understood physically by noting that lower-lying states are more tightly bound and therefore interact more strongly, increasing the probability per unit time of creating a surface electron-hole pair.

Appendices

4.A Feynman diagrams and rules

Here we give the Feynman rules for the Feynman diagrams presented in Eq. (4.24). Each diagram represents a certain number of equal terms of the series in Eq. (4.22). Firstly, there are two kinds of vertices:

$$
\boldsymbol{x}, z, \tau, \mathbf{i} \;\rightsquigarrow\; \boldsymbol{x}', \tau', \boldsymbol{s}' \;=\; \phi_{\mathrm{B};\mathbf{i}}^{*}(\boldsymbol{x}, z, \tau)\, \phi_{\mathrm{B};\mathbf{i}}(\boldsymbol{x}, z, \tau)\, \delta(\tau - \tau')
$$

$$
\times \int \mathrm{d}z'\, \pi_{s'}(z')\, V(\boldsymbol{x} - \boldsymbol{x}'; z, z'),
\tag{4.88a}
$$

$$
\otimes\!\!\rightsquigarrow\!\bullet \quad = \delta\rho_s(\boldsymbol{x}, \tau)\, \delta(\tau - \tau') \int \mathrm{d}z'\, \pi_{s'}(z')\, V(\boldsymbol{x} - \boldsymbol{x}'; z, z').
\tag{4.88b}
$$
$$
\boldsymbol{x}, \tau, \boldsymbol{s} \quad \boldsymbol{x}', \tau', \boldsymbol{s}'
$$

Secondly, the propagator is always

$$
\bullet\!-\!-\!\blacktriangleright\!-\!-\!\bullet \quad = G_{\mathrm{S}}[\rho_0]_{ss'}(\boldsymbol{x}, \tau; \boldsymbol{x}', \tau').
\tag{4.89}
$$
$$
\boldsymbol{x}, \tau, \boldsymbol{s} \quad \boldsymbol{x}', \tau', \boldsymbol{s}'
$$

Finally, there is the issue of the symmetry factor of the diagram. Consider a one-loop diagram with a total number n of external interaction vertices, that is, those in Eqs. (4.88). We need to first multiply the diagram by $\frac{1}{n}$ in accord with the prefactor in Eq. (4.22). Then we need to count the number m of ways in which the n external legs can be connected to the n fixed points of the main surface loop in order to obtain the desired diagram. The total multiplicative factor is then $\frac{m}{n}$. For instance, the diagram corresponding to $G_{\mathrm{S}}\mathcal{M}_{\mathrm{B}}G_{\mathrm{S}}\Sigma_{\mathrm{S}}G_{\mathrm{S}}\mathcal{M}_{\mathrm{B}}G_{\mathrm{S}}\Sigma_{\mathrm{S}}$ has $n = 4$ and $m = 2$, because the only possibilities to construct it are the following:
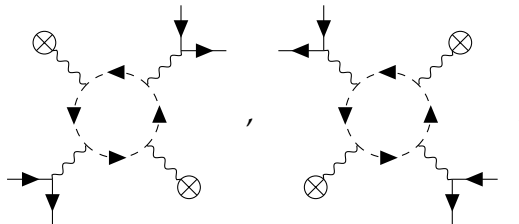
On the other hand, the diagram corresponding to $G_S\mathcal{M}_B G_S\mathcal{M}_B G_S\Sigma_S G_S\Sigma_S$ has again $n = 4$ but now $m = 4$, as the following combinations are possible:
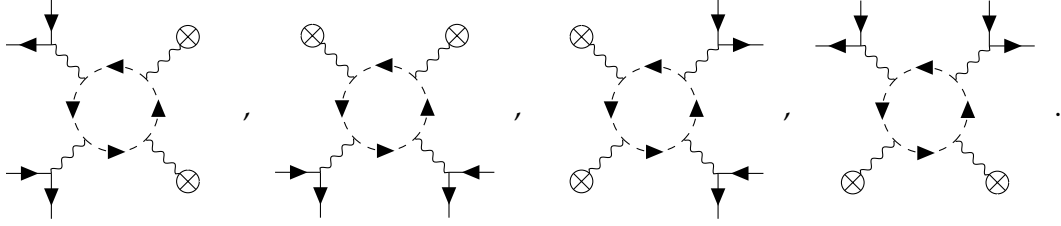
All of these diagrams arise from Eq. (4.22) and contribute equally due to the cyclicity of the trace.

The matrices $\mathcal{R}$ and $\mathcal{U}$ introduced in Eq. (4.28) can be found from all contributing Feynman diagrams by removing the external density fields $\delta\rho$. We recall that $\mathcal{R} = \sum_{n=1}^{\infty} \mathcal{R}^{(n)}$ and $\mathcal{U} = \sum_{n=0}^{\infty} \mathcal{U}^{(n)}$ and that for the two-body problem it suffices to take $\mathcal{R} = \mathcal{R}^{(1)}$ and $\mathcal{U} = \sum_{n=0}^{2} \mathcal{U}^{(n)}$. We also split $\mathcal{U}^{(2)} = \mathcal{U}^{(2,1)} + \mathcal{U}^{(2,2)}$ since there are two distinct diagrams with two external bulk-field pairs and two density fields; these are precisely the ones depicted above. We also introduce a shorthand notation $x \equiv (\boldsymbol{x}, \tau, s)$ and $G \equiv G_S[\rho_0]$, as well as $\mathcal{M}_B(x, x') \equiv \delta_{xx'}\mathcal{M}_B(x)$. Then one finds[2]

$$\mathcal{R}^{(1)}(x) = \sum_{x_1, x_2} V^{qc}(x, x_1)G(x_1, x_2)G(x_2, x_1)\mathcal{M}_B(x_2), \tag{4.90}$$

and

$$\mathcal{U}^{(0)}(x, x') = \sum_{x_1, x_2} V^{qc}(x, x_1)G(x_1, x_2)G(x_2, x_1)V^{qc}(x_2, x'), \tag{4.91a}$$

$$\mathcal{U}^{(1)}(x, x') = 2\sum_{x_1, x_2, x_3} V^{qc}(x, x_1)V^{qc}(x', x_2)G(x_1, x_2)G(x_2, x_3)G(x_3, x_1)\mathcal{M}_B(x_3), \tag{4.91b}$$

$$\mathcal{U}^{(2,1)}(x, x') = \sum_{x_1, x_2, x_3, x_4} V^{qc}(x, x_1)V^{qc}(x', x_3)G(x_1, x_2)G(x_2, x_3)G(x_3, x_4)G(x_4, x_1)\mathcal{M}_B(x_2)\mathcal{M}_B(x_4),$$
$$\tag{4.91c}$$

$$\mathcal{U}^{(2,2)}(x, x') = 2\sum_{x_1, x_2, x_3, x_4} V^{qc}(x, x_1)V^{qc}(x', x_2)G(x_1, x_2)G(x_2, x_3)G(x_3, x_4)G(x_4, x_1)\mathcal{M}_B(x_3)\mathcal{M}_B(x_4).$$
$$\tag{4.91d}$$

Note that the combinatorial factor in the diagrams contributing to $\mathcal{U}$ must be multiplied by an additional factor of 2 to compensate for the factor of $\frac{1}{2}$ that we have added in the last term of Eq. (4.28).

---

[2]We remind the reader that the interaction is instantaneous in imaginary time, so in all of these expressions it is understood that $V^{qc}(x, x') = V^{qc}_{ss'}(\boldsymbol{x} - \boldsymbol{x}')\delta(\tau - \tau')$.

5
Summary and outlook

In this work, we have studied the properties of Wannier excitons in the bulk of thin Bi_2Se_3 nanosheets. The Coulomb interaction is heavily modified by both the thin-slab geometry and the topological band structure of the material. As a consequence of this, we have found that s-wave excitons consist of a nonchiral doublet with quadratic dispersion at low momenta, and a chiral doublet with one nonanalytic linear mode and one quadratic mode. Our results show that in principle the topology of the conduction and valence bands is inherited by the chiral excitons, as the topologically nontrivial windings of the underlying electron and hole states add up to give a total winding number of two around $Q = 0$. This winding manifests itself as a nonzero Chern number once a small gap is opened between the two zero-momentum states of the chiral doublet. Based on the topological properties of the effective model for the ground-state chiral doublet, we have also predicted the existence of chiral excitonic edge states. However, we have not yet succeeded in the explicit determination of their wave functions and energies, an issue that must thus be addressed in future work on the topic. Additionally, we have investigated some optical properties of the excitons at zero momentum. All excitons whose constituent electron and hole have opposite spin-orbit parities are found to be dark under circular polarization. For equal parities, s-wave and d-wave excitons are selectively bright under left- or right-circular polarization.

We have also analyzed the effects of the coherent surface states on the bulk electron-hole interaction, as they are expected to provide the main contribution to this phenomenon. A formula for the screened interaction in the RPA scheme has been derived in the limit of perfectly localized surface states and used to describe Thomas-Fermi screening of some common interaction potentials. This formula in principle allows for a detailed microscopic description of the interaction potential as a function of the vertical position inside the nanosheet, as exemplified with the Rytova-Keldysh potential. The surface plasmons have been studied in the RPA as well, and we have found an optical branch and an acoustic branch. Due to these plasmons, decay of bulk excitons into surface electron-hole pairs is in principle a possible mechanism but its effect is negligible in the static approximation.

Since our work does not impose many restrictions on the model parameters, we expect that it remains valid for other similar topological materials such as Bi_2Te_3 and Sb_2Se_3 by using the appropriate parameter values.

Despite the theoretical nature of the present work, many aspects are experimentally relevant and could lead to a variety of applications outlined below. In particular, it would be of interest to resolve the linear dispersion in the chiral doublet. This may be achieved by means of terahertz spectroscopy, as the details of the dispersion affect the chemical equilibrium between excitons and free charges in such pump-probe experiments [124–126]. The response to light of chiral excitons would also be interesting to consider in order to probe our derived selection rules and verify the different exciton types.

The work presented in this thesis concerns the treatment of *free* excitons inside Bi_2Se_3. Knowledge of their behavior is undoubtedly important on its own, as their presence or absence is crucial in optoelectronic devices such as lasers [127, 128], light-emitting diodes [129, 130] and photovoltaic cells [131, 132]. Even so, however, our results are merely the first building block of a much bigger picture, because interaction effects of bulk excitons with other (quasi)particles have not been studied here. Both from a fundamental and from a practical point of view, the physics of such interactions are the richest and most exciting aspect of the study of excitons in semiconductors.

For instance, interactions between excitons and photons give rise to quasiparticles known as *exciton-polaritons*. Such polaritonic modes have previously been achieved at room temperature by Vanmaekelbergh *et al.* [133] and can be used for nanophotonic circuitry and waveguiding with minimal losses [134]. Recent experimental observations strongly suggest that the topological character of excitons is preserved in their recombination [135], allowing for the conversion of spin into circular rotation of the emitted photon. Thus, exciton-polaritons enable waveguide transfer of robust photonic states, which can then carry information over macroscopic distances with minimal decoherence [136]. They may also undergo Bose-Einstein condensation [137] and behave like a so-called *superfluid*, which flows around obstacles without dissipation [138]. Promising advances have been made toward the achievement of such a condensate at room temperature, which would allow for the development of quasi-lossless quantum devices [139].

On the other hand, coupling between excitons and surface plasmons is another interesting topic the basics of which have already been addressed in the previous chapter. Exciton-plasmon interactions may allow for the design of specific absorption and emission properties, enabling increased control of the optical behavior and energy flows almost all the way down to atomic length and time scales [140]. This could make them suitable for efficient energy-conversion, absorption and transfer applications [140–143]. They also enable ultrafast manipulation of spin states [144] and exciton-exciton hybridization offering room-temperature quantum-information transport and spin entanglement [145].

Exciton-exciton interactions are also of interest. They have been explored to represent qubits, the degrees of freedom of a quantum computer [146, 147], although in this case the particles are confined in quantum dots and the excitons are thus of the Frenkel type. Moreover, since excitons are essentially bosonic particles, interactions between them may cause them to undergo Bose-Einstein condensation [148–150]. We have found that excitons in Bi_2Se_3 are indirect, which is desirable due to a reduced recombination rate and leads to the exciting possibility of forming a *supersolid*, that is, a superfluid with periodic density order [151, 152].

All in all, the many-body physics of topological excitons and their interactions are very rich and offer a vast number of possibilities for exciting new research in condensed-matter physics. The free-exciton picture, which is now well understood, can be systematically refined and expanded with the inclusion of such effects. Our work paves the way for the analysis of these and other phenomena, which may lead to exotic phases of matter and exciting optoelectronics and quantum-information applications.

Bibliography

[1]   H.-B. Braun. "Topological effects in nanomagnetism: from superparamagnetism to chiral quantum solitons". *Advances in Physics* 61.1 (2012), pp. 1–116. DOI: 10.1080/00018732.2012.663070.

[2]   J. Zang, V. Cros, and A. Hoffmann. *Topology in magnetism*. Vol. 192. Springer, 2018. DOI: 10.1007/978-3-319-97334-0.

[3]   M. Nakahara. *Geometry, topology and physics*. CRC press, 2018. DOI: 10.1201/9781315275826.

[4]   B. F. Schutz. *Geometrical methods of mathematical physics*. Cambridge University Press, 1980. DOI: 10.1017/CBO9781139171540.

[5]   C. Kittel and C.-Y. Fong. *Quantum theory of solids*. Vol. 5. Wiley New York, 1963. ISBN: 978-0-471-62412-7.

[6]   N. W. Ashcroft, N. D. Mermin, et al. *Solid state physics*. Saunders College Publishing, 1976. DOI: 10.1119/1.11117.

[7]   M. Z. Hasan and C. L. Kane. "Colloquium: topological insulators". *Reviews of Modern Physics* 82.4 (2010), p. 3045. DOI: 10.1103/RevModPhys.82.3045.

[8]   X.-L. Qi and S.-C. Zhang. "Topological insulators and superconductors". *Reviews of Modern Physics* 83.4 (2011), p. 1057. DOI: 10.1103/RevModPhys.83.1057.

[9]   A. W. Ludwig. "Topological phases: classification of topological insulators and superconductors of non-interacting fermions, and beyond". *Physica Scripta* 2016.T168 (2015), p. 014001. DOI: 10.1088/0031-8949/2015/T168/014001.

[10] C.-K. Chiu et al. "Classification of topological quantum matter with symmetries". *Reviews of Modern Physics* 88.3 (2016), p. 035005. DOI: 10.1103/RevModPhys.88.035005.

[11] A. P. Schnyder et al. "Classification of topological insulators and superconductors in three spatial dimensions". *Physical Review B* 78.19 (2008), p. 195125. DOI: 10.1103/PhysRevB.78.195125.

[12] A. P. Schnyder et al. "Classification of topological insulators and superconductors". *AIP Conference Proceedings*. Vol. 1134. 1. American Institute of Physics. 2009, pp. 10–21. DOI: 10.1063/1.3149481.

[13] S. Ryu et al. "Topological insulators and superconductors: tenfold way and dimensional hierarchy". *New Journal of Physics* 12.6 (2010), p. 065010. DOI: 10.1088/1367-2630/12/6/065010.

[14] A. Kitaev. "Periodic table for topological insulators and superconductors". *AIP Conference Proceedings*. Vol. 1134. 1. American Institute of Physics. 2009, pp. 22–30. DOI: 10.1063/1.3149495.

[15] S. Ryu, J. E. Moore, and A. W. Ludwig. "Electromagnetic and gravitational responses and anomalies in topological insulators and superconductors". *Physical Review B* 85.4 (2012), p. 045104. DOI: 10.1103/PhysRevB.85.045104.

[16] M. E. Cage et al. *The quantum Hall effect*. Springer Science & Business Media, 1990. DOI: `10.1007/978-1-4612-3350-3`.

[17] D. Yoshioka. *The quantum Hall effect*. Vol. 133. Springer Science & Business Media, 2002. DOI: `10.1007/978-3-662-05016-3`.

[18] E. H. Hall et al. "On a new action of the magnet on electric currents". *American Journal of Mathematics* 2.3 (1879), pp. 287–292. DOI: `10.1038/021361a0`.

[19] M. Suddards et al. "Scanning capacitance imaging of compressible and incompressible quantum Hall effect edge strips". *New Journal of Physics* 14.8 (2012), p. 083015. DOI: `10.1088/1367-2630/14/8/083015`.

[20] K. v. Klitzing, G. Dorda, and M. Pepper. "New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance". *Physical Review Letters* 45.6 (1980), p. 494. DOI: `10.1103/PhysRevLett.45.494`.

[21] K. Von Klitzing. "The quantized Hall effect". *Reviews of Modern Physics* 58.3 (1986), p. 519. DOI: `10.1103/RevModPhys.58.519`.

[22] K. Von Klitzing. "Developments in the quantum Hall effect". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 363.1834 (2005), pp. 2203–2219. DOI: `10.1098/rsta.2005.1640`.

[23] R. B. Laughlin. "Quantized Hall conductivity in two dimensions". *Physical Review B* 23.10 (1981), p. 5632. DOI: `10.1103/PhysRevB.23.5632`.

[24] D. J. Thouless et al. "Quantized Hall conductance in a two-dimensional periodic potential". *Physical Review Letters* 49.6 (1982), p. 405. DOI: `10.1103/PhysRevLett.49.405`.

[25] I. Dana and J. Zak. "Adams representation and localization in a magnetic field". *Physical Review B* 28.2 (1983), p. 811. DOI: `10.1103/PhysRevB.28.811`.

[26] I. Dana, Y. Avron, and J. Zak. "Quantised Hall conductance in a perfect crystal". *Journal of Physics C: Solid State Physics* 18.22 (1985), p. L679. DOI: `10.1088/0022-3719/18/22/004/meta`.

[27] F. D. M. Haldane. "Model for a quantum Hall effect without Landau levels: Condensed-matter realization of the "parity anomaly"". *Physical Review Letters* 61.18 (1988), p. 2015. DOI: `10.1103/PhysRevLett.61.2015`.

[28] C. L. Kane and E. J. Mele. "Quantum spin Hall effect in graphene". *Physical Review Letters* 95.22 (2005), p. 226801. DOI: `10.1103/PhysRevLett.95.226801`.

[29] C. L. Kane and E. J. Mele. "$\mathbb{Z}_2$ topological order and the quantum spin Hall effect". *Physical Review Letters* 95.14 (2005), p. 146802. DOI: `10.1103/PhysRevLett.95.146802`.

[30] L. Fu and C. L. Kane. "Topological insulators with inversion symmetry". *Physical Review B* 76.4 (2007), p. 045302. DOI: `10.1103/PhysRevB.76.045302`.

[31] H. A. Kramers. "Théorie générale de la rotation paramagnétique dans les cristaux". *Proceedings of the Royal Academy of Sciences at Amsterdam* 33.6 (1930).

[32] M. J. Klein. "On a degeneracy theorem of Kramers". *American Journal of Physics* 20.2 (1952), pp. 65–71. DOI: `10.1119/1.1933118`.

[33] L. Fu, C. L. Kane, and E. J. Mele. "Topological insulators in three dimensions". *Physical Review Letters* 98.10 (2007), p. 106803. DOI: `10.1103/PhysRevLett.98.106803`.

[34] J. C. Teo, L. Fu, and C. Kane. "Surface states and topological invariants in three-dimensional topological insulators: Application to $Bi_{1-x}Sb_x$". *Physical Review B* 78.4 (2008), p. 045426. DOI: `10.1103/PhysRevB.78.045426`.

[35] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang. "Quantum spin Hall effect and topological phase transition in HgTe quantum wells". *Science* 314.5806 (2006), pp. 1757–1761. DOI: `10.1126/science.1133734`.

[36] M. König et al. "The quantum spin Hall effect: theory and experiment". *Journal of the Physical Society of Japan* 77.3 (2008), p. 031007. DOI: `10.1143/JPSJ.77.031007`.

[37] R. S. Mong and V. Shivamoggi. "Edge states and the bulk-boundary correspondence in Dirac Hamiltonians". *Physical Review B* 83.12 (2011), p. 125109. DOI: `10.1103/PhysRevB.83.125109`.

[38] A. Alase. *Boundary Physics and Bulk-Boundary Correspondence in Topological Phases of Matter*. Dartmouth College, 2019. DOI: `10.1007/978-3-030-31960-1`.

[39] A. Bansil, H. Lin, and T. Das. "Colloquium: Topological band theory". *Reviews of Modern Physics* 88.2 (2016), p. 021004. DOI: `10.1103/RevModPhys.88.021004`.

[40] X.-L. Qi, T. L. Hughes, and S.-C. Zhang. "Topological field theory of time-reversal invariant insulators". *Physical Review B* 78.19 (2008), p. 195424. DOI: `10.1103/PhysRevB.78.195424`.

[41] S. Ryu. "Interacting topological phases and quantum anomalies". *Physica Scripta* 2015.T164 (2015), p. 014009. DOI: `10.1088/0031-8949/2015/T164/014009`.

[42] E. Witten. "Fermion path integrals and topological phases". *Reviews of Modern Physics* 88.3 (2016), p. 035001. DOI: `10.1103/RevModPhys.88.035001`.

[43] C.-T. Hsieh, G. Y. Cho, and S. Ryu. "Global anomalies on the surface of fermionic symmetry-protected topological phases in (3+ 1) dimensions". *Physical Review B* 93.7 (2016), p. 075135. DOI: `10.1103/PhysRevB.93.075135`.

[44] C. Wang and T. Senthil. "Interacting fermionic topological insulators/superconductors in three dimensions". *Physical Review B* 89.19 (2014), p. 195124. DOI: `10.1103/PhysRevB.89.195124`.

[45] C. Xu and J. E. Moore. "Stability of the quantum spin Hall effect: Effects of interactions, disorder, and Z 2 topology". *Physical Review B* 73.4 (2006), p. 045322. DOI: `10.1103/PhysRevB.73.045322`.

[46] C. Wu, B. A. Bernevig, and S.-C. Zhang. "Helical liquid and the edge of quantum spin Hall systems". *Physical Review Letters* 96.10 (2006), p. 106401. DOI: `10.1103/PhysRevLett.96.106401`.

[47] S.-S. Lee and S. Ryu. "Many-body generalization of the $\mathbb{Z}_2$ topological invariant for the quantum spin Hall effect". *Physical Review Letters* 100.18 (2008), p. 186807. DOI: `10.1103/PhysRevLett.100.186807`.

[48] M. V. Berry. "Quantal phase factors accompanying adiabatic changes". *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences* 392.1802 (1984), pp. 45–57. DOI: `10.1098/rspa.1984.0023`.

[49] D. Xiao, M.-C. Chang, and Q. Niu. "Berry phase effects on electronic properties". *Reviews of Modern Physics* 82.3 (2010), p. 1959. DOI: `10.1103/RevModPhys.82.1959`.

[50] D. Vanderbilt. *Berry Phases in Electronic Structure Theory: Electric Polarization, Orbital Magnetization and Topological Insulators*. Cambridge University Press, 2018. DOI: `10.1017/9781316662205`.

[51] T. Kato. "On the adiabatic theorem of quantum mechanics". *Journal of the Physical Society of Japan* 5.6 (1950), pp. 435–439. DOI: `10.1143/JPSJ.5.435`.

[52] F. Wilczek and A. Zee. "Appearance of gauge structure in simple dynamical systems". *Physical Review Letters* 52.24 (1984), p. 2111. DOI: `10.1103/PhysRevLett.52.2111`.

[53] H. Zhang et al. "Topological insulators in Bi_2Se_3, Bi_2Te_3 and Sb_2Te_3 with a single Dirac cone on the surface". *Nature Physics* 5.6 (2009), pp. 438–442. DOI: `10.1038/NPHYS1270`.

[54] C.-X. Liu et al. "Model Hamiltonian for topological insulators". *Physical Review B* 82.4 (2010), p. 045122. DOI: `10.1103/PhysRevB.82.045122`.

[55] R. S. Knox. *Solid State Physics: Supplement*. 5. Academic Press, 1963.

[56] W. Liang. "Excitons". *Physics Education* 5.4 (1970), p. 226. DOI: `10.1088/0031-9120/5/4/003`.

[57] J. I. Pankove. *Optical processes in semiconductors*. Courier Corporation, 1975. DOI: `10.1016/0022-2313(72)90050-6`.

[58] J. Singh. *Excitation energy transfer processes in condensed matter: theory and applications*. Springer Science & Business Media, 2013. DOI: `10.1007/978-1-4899-0996-1`.

[59] M. Combescot and S.-Y. Shiau. *Excitons and Cooper pairs: two composite bosons in many-body physics*. Oxford University Press, 2015. DOI: `10.1093/acprof:oso/9780198753735.001.0001`.

[60] J. Frenkel. "On the transformation of light into heat in solids. I". *Physical Review* 37.1 (1931), p. 17. DOI: `10.1103/PhysRev.37.17`.

[61] G. H. Wannier. "The structure of electronic excitation levels in insulating crystals". *Physical Review* 52.3 (1937), p. 191. DOI: `10.1103/PhysRev.52.191`.

[62]   E. U. Condon and G. H. Shortley. *The theory of atomic spectra*. Cambridge University Press, 1935. ISBN: 978-0-521-09209-8.

[63]   R. Elliott. "Intensity of optical absorption by excitons". *Physical Review* 108.6 (1957), p. 1384. DOI: 10.1103/PhysRev.108.1384.

[64]   G. Dresselhaus. "Effective mass approximation for excitons". *Journal of Physics and Chemistry of Solids* 1.1-2 (1956), pp. 14–22. DOI: 10.1016/0022-3697(56)90004-X.

[65]   N. Lipari and M. Altarelli. "Theory of indirect excitons in semiconductors". *Physical Review B* 15.10 (1977), p. 4883. DOI: 10.1103/PhysRevB.15.4883.

[66]   G. Macfarlane et al. "Exciton and phonon effects in the absorption spectra of germanium and silicon". *Journal of Physics and Chemistry of Solids* 8 (1959), pp. 388–392. DOI: 10.1016/0022-3697(59)90372-5.

[67]   E. Calman et al. "Indirect excitons in van der Waals heterostructures at room temperature". *Nature Communications* 9.1 (2018), pp. 1–5. DOI: 10.1038/s41467-018-04293-7.

[68]   J. R. Madureira et al. "Spatially indirect excitons in type-II quantum dots". *Applied Physics Letters* 90.21 (2007), p. 212105. DOI: 10.1063/1.2741601.

[69]   Y. Peter and M. Cardona. *Fundamentals of semiconductors: physics and materials properties*. Springer Science & Business Media, 2010. DOI: 10.1007/978-3-642-00710-1.

[70]   X.-L. Qi and S.-C. Zhang. "The quantum spin Hall effect and topological insulators". *Physics Today* 63.1 (2010), pp. 33–38. DOI: 10.1063/1.3293411.

[71]   H.-Z. Lu et al. "Massive Dirac fermions and spin physics in an ultrathin film of topological insulator". *Physical Review B* 81.11 (2010), p. 115407. DOI: 10.1103/PhysRevB.81.115407.

[72]   W.-Y. Shan, H.-Z. Lu, and S.-Q. Shen. "Effective continuous model for surface states and thin films of three-dimensional topological insulators". *New Journal of Physics* 12.4 (2010), p. 043048. DOI: 10.1088/1367-2630/12/4/043048.

[73]   J. Linder, T. Yokoyama, and A. Sudbø. "Anomalous finite size effects on surface states in the topological insulator Bi_2Se_3". *Physical Review B* 80.20 (2009), p. 205401. DOI: 10.1103/PhysRevB.80.205401.

[74]   P. Larson et al. "Electronic structure of Bi_2X_3 (X = S, Se, T) compounds: Comparison of theoretical calculations with photoemission studies". *Physical Review B* 65.8 (2002), p. 085108. DOI: 10.1103/PhysRevB.65.085108.

[75]   O. V. Yazyev, J. E. Moore, and S. G. Louie. "Spin polarization and transport of surface states in the topological insulators Bi_2Se_3 and Bi_2Te_3 from first principles". *Physical Review Letters* 105.26 (2010), p. 266806. DOI: 10.1103/PhysRevLett.105.266806.

[76]   J. Black et al. "Electrical and optical properties of some $M_2^{V–B}N_3^{VI–B}$ semiconductors". *Journal of Physics and Chemistry of Solids* 2.3 (1957), pp. 240–251. DOI: 10.1016/0022-3697(57)90090-2.

[77]   C. M. Acosta et al. "Tight-binding model for the band dispersion in rhombohedral topological insulators over the whole Brillouin zone". *Physical Review B* 98.3 (2018), p. 035106. DOI: 10.1103/PhysRevB.98.035106.

[78]   B. Zhou et al. "Finite size effects on helical edge states in a quantum spin-Hall system". *Physical Review Letters* 101.24 (2008), p. 246807. DOI: 10.1103/PhysRevLett.101.246807.

[79]   C.-X. Liu et al. "Oscillatory crossover from two-dimensional to three-dimensional topological insulators". *Physical Review B* 81.4 (2010), p. 041307. DOI: 10.1103/PhysRevB.81.041307.

[80]   Y. Zhang et al. "Crossover of the three-dimensional topological insulator Bi_2Se_3 to the two-dimensional limit". *Nature Physics* 6.8 (2010), pp. 584–588. DOI: 10.1038/NPHYS1689.

[81]   Z. Wang et al. "Dimensional crossover and topological nature of the thin films of a three-dimensional topological insulator by band gap engineering". *Nano Letters* 19.7 (2019), pp. 4627–4633. DOI: 10.1021/acs.nanolett.9b01641.

[82]   M. Hasan and M. Majidi. "Electronic structure of 9 quintuple layers Bi2Se3 within Density Functional Theory". *IOP Conference Series: Materials Science and Engineering*. Vol. 902. 1. IOP Publishing. 2020, p. 012061. DOI: 10.1088/1757-899X/902/1/012061.

[83] H. Okamoto. "The Bi-Se (bismuth-selenium) system". *Journal of Phase Equilibria* 15.2 (1994), pp. 195–201. DOI: 10.1007/BF02646366.

[84] H. Lind, S. Lidin, and U. Häussermann. "Structure and bonding properties of $(Bi_2Se_3)_m(Bi_2)_n$ stacks by first-principles density functional theory". *Physical Review B* 72.18 (2005), p. 184101. DOI: 10.1103/PhysRevB.72.184101.

[85] G. Zhang et al. "Quintuple-layer epitaxy of thin films of topological insulator Bi_2Se_2". *Applied Physics Letters* 95.5 (2009), p. 053114. DOI: 10.1063/1.3200237.

[86] A. Taskin et al. "Manifestation of topological protection in transport properties of epitaxial Bi_2Se_3 thin films". *Physical Review Letters* 109.6 (2012), p. 066803. DOI: 10.1103/PhysRevLett.109.066803.

[87] G. Bihlmayer, O. Rader, and R. Winkler. "Focus on the Rashba effect". *New Journal of Physics* 17.5 (2015), p. 050202. DOI: 10.1088/1367-2630/17/5/050202.

[88] P. Li, I. Appelbaum, et al. "Illuminating "spin-polarized" Bloch wave-function projection from degenerate bands in decomposable centrosymmetric lattices". *Physical Review B* 97.12 (2018), p. 125434. DOI: 10.1103/PhysRevB.97.125434.

[89] R. Moessner and J. E. Moore. *Topological phases of matter*. Cambridge University Press, 2021. DOI: 10.1017/9781316226308.

[90] H. T. Stoof, K. B. Gubbels, and D. B. Dickerscheid. *Ultracold quantum fields*. Vol. 1. Springer, 2009. DOI: 10.1007/978-1-4020-8763-9.

[91] L. Keldysh. "Coulomb interaction in thin semiconductor and semimetal films". *Soviet Journal of Experimental and Theoretical Physics Letters* 29 (1979), p. 658.

[92] D. Van Tuan, M. Yang, and H. Dery. "Coulomb interaction in monolayer transition-metal dichalcogenides". *Physical Review B* 98.12 (2018), p. 125308. DOI: 10.1103/PhysRevB.98.125308.

[93] N. Leppenen, L. Golub, and E. Ivchenko. "Exciton oscillator strength in two-dimensional Dirac materials". *Physical Review B* 102.15 (2020), p. 155305. DOI: 10.1103/PhysRevB.102.155305.

[94] Z. Feng et al. "Dielectric Properties of Two-Dimensional Bi2Se3 Hexagonal Nanoplates Modified PVDF Nanocomposites". *Advances in Polymer Technology* 2019 (2019). DOI: 10.1155/2019/8720678.

[95] T. Cao and S. Wang. "Topological insulator metamaterials with tunable negative refractive index in the optical region". *Nanoscale Research Letters* 8.1 (2013), pp. 1–8. DOI: 10.1186/1556-276X-8-526.

[96] M. Fang et al. "Layer-dependent dielectric permittivity of topological insulator Bi_2Se_3 thin films". *Applied Surface Science* 509 (2020), p. 144822. DOI: 10.1016/j.apsusc.2019.144822.

[97] F. G. Flórez, L. D. Siebbeles, and H. Stoof. "Effects of material thickness and surrounding dielectric medium on Coulomb interactions and two-dimensional excitons". *Physical Review B* 102.12 (2020), p. 125303. DOI: 10.1103/PhysRevB.102.125303.

[98] M. V. Sadovskii. *Diagrammatics: Lectures on Selected Problems in Condensed Matter Theory*. World Scientific, 2006. DOI: 10.1142/11605.

[99] W. A. Karr, C. R. Jamell, and Y. N. Joglekar. "Numerical approach to the Schrödinger equation in momentum space". *American Journal of Physics* 78.4 (2010), pp. 407–411. DOI: 10.1119/1.3272021.

[100] D. Y. Qiu, T. Cao, and S. G. Louie. "Nonanalyticity, valley quantum phases, and lightlike exciton dispersion in monolayer transition metal dichalcogenides: Theory and first-principles calculations". *Physical Review Letters* 115.17 (2015), p. 176801. DOI: 10.1103/PhysRevLett.115.176801.

[101] D. Parfitt and M. Portnoi. "The two-dimensional hydrogen atom revisited". *Journal of Mathematical Physics* 43.10 (2002), pp. 4681–4691. DOI: 10.1063/1.1503868.

[102] J. Zhou et al. "Berry phase modification to the energy spectrum of excitons". *Physical Review Letters* 115.16 (2015), p. 166803. DOI: 10.1103/PhysRevLett.115.166803.

[103] X. Zhang, W.-Y. Shan, and D. Xiao. "Optical selection rule of excitons in gapped chiral fermion systems". *Physical Review Letters* 120.7 (2018), p. 077401. DOI: 10.1103/PhysRevLett.120.077401.

[104] L. C. Andreani and F. Bassani. "Exchange interaction and polariton effects in quantum-well excitons". *Physical Review B* 41.11 (1990), p. 7536. DOI: 10.1103/PhysRevB.41.7536.

[105] H. Yu et al. "Dirac cones and Dirac saddle points of bright excitons in monolayer transition metal dichalcogenides". *Nature Communications* 5.1 (2014), pp. 1–7. DOI: 10.1038/ncomms4876.

[106] F. Wu, F. Qu, and A. H. MacDonald. "Exciton band structure of monolayer MoS$_2$". *Physical Review B* 91.7 (2015), p. 075310. DOI: 10.1103/PhysRevB.91.075310.

[107] M.-C. Chang and Q. Niu. "Berry phase, hyperorbits, and the Hofstadter spectrum". *Physical Review Letters* 75.7 (1995), p. 1348. DOI: 10.1103/PhysRevLett.75.1348.

[108] M.-C. Chang and Q. Niu. "Berry phase, hyperorbits, and the Hofstadter spectrum: Semiclassical dynamics in magnetic Bloch bands". *Physical Review B* 53.11 (1996), p. 7010. DOI: 10.1103/PhysRevB.53.7010.

[109] G. Sundaram and Q. Niu. "Wave-packet dynamics in slowly perturbed crystals: Gradient corrections and Berry-phase effects". *Physical Review B* 59.23 (1999), p. 14915. DOI: 10.1103/PhysRevB.59.14915.

[110] T. Cao, M. Wu, and S. G. Louie. "Unifying optical selection rules for excitons in two dimensions: Band topology and winding numbers". *Physical Review Letters* 120.8 (2018), p. 087402. DOI: 10.1103/PhysRevLett.120.087402.

[111] T. G. Pedersen, K. Pedersen, and T. B. Kriestensen. "Optical matrix elements in tight-binding calculations". *Physical Review B* 63.20 (2001), p. 201101. DOI: 10.1103/PhysRevB.63.201101.

[112] Z. Gong et al. "Chiral topological excitons in the monolayer transition metal dichalcogenides". *Scientific Reports* 7.1 (2017), pp. 1–8. DOI: 10.1038/srep42390.

[113] A. Altland and B. D. Simons. *Condensed matter field theory*. Cambridge University Press, 2010. DOI: 10.1017/CBO9780511789984.

[114] A. Ceferino et al. "Crossover from weakly indirect to direct excitons in atomically thin films of InSe". *Physical Review B* 101.24 (2020), p. 245432. DOI: 10.1103/PhysRevB.101.245432.

[115] R. D. Mattuck. *A guide to Feynman diagrams in the many-body problem*. Courier Corporation, 1992. DOI: 10.1063/1.3035595.

[116] K. W.-K. Shung. "Dielectric function and plasmon structure of stage-1 intercalated graphite". *Physical Review B* 34.2 (1986), p. 979. DOI: 10.1103/PhysRevB.34.979.

[117] B. Wunsch et al. "Dynamical polarization of graphene at finite doping". *New Journal of Physics* 8.12 (2006), p. 318. DOI: 10.1088/1367-2630/8/12/318.

[118] E. Hwang and S. D. Sarma. "Dielectric function, screening, and plasmons in two-dimensional graphene". *Physical Review B* 75.20 (2007), p. 205418. DOI: 10.1103/PhysRevB.75.205418.

[119] F. Stern. "Polarizability of a two-dimensional electron gas". *Physical Review Letters* 18.14 (1967), p. 546. DOI: 10.1103/PhysRevLett.18.546.

[120] E. O. Kane. "Thomas-Fermi approach to impure semiconductor band structure". *Physical Review* 131.1 (1963), p. 79. DOI: 10.1103/PhysRev.131.79.

[121] R. Resta. "Thomas-Fermi dielectric screening in semiconductors". *Physical Review B* 16.6 (1977), p. 2717. DOI: 10.1103/PhysRevB.16.2717.

[122] B. Firey and N. Ashcroft. "Thermodynamics of Thomas-Fermi screened Coulomb systems". *Physical Review A* 15.5 (1977), p. 2072. DOI: 10.1103/PhysRevA.15.2072.

[123] G. Gumbs et al. "Plasmon excitations of multi-layer graphene on a conducting substrate". *Scientific Reports* 6.1 (2016), pp. 1–11. DOI: 10.1038/srep21063.

[124] J. Lauth, S. Kinge, and L. D. Siebbeles. "Ultrafast Transient Absorption and Terahertz Spectroscopy as Tools to Probe Photoexcited States and Dynamics in Colloidal 2D Nanostructures". *Zeitschrift für Physikalische Chemie* 231.1 (2017), pp. 107–119. DOI: 10.1515/zpch-2016-0911.

[125] F. G. Flórez et al. "Explaining observed stability of excitons in highly excited CdSe nanoplatelets". *Physical Review B* 100.24 (2019), p. 245302. DOI: 10.1103/PhysRevB.100.245302.

[126] J. Lauth et al. "Photoexcitation of PbS nanosheets leads to highly mobile charge carriers and stable excitons". *Nanoscale* 11.44 (2019), pp. 21569–21576. DOI: 10.1039/C9NR07927K.

[127] W. Wen, L. Wu, and T. Yu. "Excitonic lasers in atomically thin 2D semiconductors". *ACS Materials Letters* 2.10 (2020), pp. 1328–1342. DOI: 10.1021/acsmaterialslett.0c00277.

[128] D. Vanmaekelbergh and L. K. Van Vugt. "ZnO nanowire lasers". *Nanoscale* 3.7 (2011), pp. 2783–2800. DOI: 10.1039/c1nr00013f.

[129] S. Tsintzos et al. "A GaAs polariton light-emitting diode operating near room temperature". *Nature* 453.7193 (2008), pp. 372–375. DOI: 10.1038/nature06979.

[130] H. Uoyama et al. "Highly efficient organic light-emitting diodes from delayed fluorescence". *Nature* 492.7428 (2012), pp. 234–238. DOI: 10.1038/nature11687.

[131] J. M. Luther and J. C. Johnson. "An exciting boost for solar cells". *Nature* 571.7763 (2019), pp. 38–39. DOI: 10.1038/d41586-019-02014-8.

[132] D. N. Congreve et al. "External quantum efficiency above 100% in a singlet-exciton-fission–based organic photovoltaic cell". *Science* 340.6130 (2013), pp. 334–337. DOI: 10.1126/science.1232994.

[133] L. K. van Vugt et al. "Exciton polaritons confined in a ZnO nanowire cavity". *Physical Review Letters* 97.14 (2006), p. 147401. DOI: 10.1103/PhysRevLett.97.147401.

[134] Z. Li et al. "High-Quality All-Inorganic Perovskite $CsPbBr_3$ Microsheet Crystals as Low-Loss Subwavelength Exciton–Polariton Waveguides". *Nano Letters* 21.4 (2021), pp. 1822–1830. DOI: 10.1021/acs.nanolett.0c04908.

[135] H.-H. Kung et al. "Observation of chiral surface excitons in a topological insulator Bi_2Se_3". *Proceedings of the National Academy of Sciences* 116.10 (2019), pp. 4006–4011. DOI: 10.1073/pnas.1813514116.

[136] M. Segev and M. A. Bandres. "Topological photonics: Where do we go from here?" *Nanophotonics* 10.1 (2021), pp. 425–434. DOI: 10.1515/nanoph-2020-0441.

[137] J. Kasprzak et al. "Bose–Einstein condensation of exciton polaritons". *Nature* 443.7110 (2006), pp. 409–414. DOI: 10.1038/nature05131.

[138] N. Bogoliubov. "On the theory of superfluidity". *Journal of Physics* 11.1 (1947), p. 23.

[139] H. Deng, H. Haug, and Y. Yamamoto. "Exciton-polariton Bose-Einstein condensation". *Reviews of Modern Physics* 82.2 (2010), p. 1489. DOI: 10.1103/RevModPhys.82.1489.

[140] M. Achermann. "Exciton-plasmon interactions in metal-semiconductor nanostructures". *The Journal of Physical Chemistry Letters* 1.19 (2010), pp. 2837–2843. DOI: 10.1021/jz101102e.

[141] P. V. Kamat and G. V. Hartland. "Plasmons for energy conversion". *ACS Energy Letters* 3.6 (2018), pp. 1467–1469. DOI: 10.1021/acsenergylett.8b00721.

[142] S. A. Maier et al. "Local detection of electromagnetic energy transport below the diffraction limit in metal nanoparticle plasmon waveguides". *Nature Materials* 2.4 (2003), pp. 229–232. DOI: 10.1038/nmat852.

[143] M. El Kabbash et al. "Plasmon-exciton resonant energy transfer: across scales hybrid systems". *Journal of Nanomaterials* 2016 (2016). DOI: 10.1155/2016/4819040.

[144] P. Yin and P. V. Radovanovic. "Magnetoplasmon resonances in semiconductor nanocrystals: Potential for a new information technology platform". *ChemSusChem* 13.18 (2020), pp. 4885–4893. DOI: 10.1002/cssc.202001468.

[145] A. H. Rose et al. "Plasmon-mediated coherent superposition of discrete excitons under strong exciton-plasmon coupling in few-layer MoS_2 at room temperature". *ACS Photonics* 7.5 (2020), pp. 1129–1134. DOI: 10.1021/acsphotonics.0c00233.

[146] F. Troiani, U. Hohenester, and E. Molinari. "Exploiting exciton-exciton interactions in semiconductor quantum dots for quantum-information processing". *Physical Review B* 62.4 (2000), R2263. DOI: 10.1103/PhysRevB.62.R2263.

[147] H. J. Krenner et al. "Recent advances in exciton-based quantum information processing in quantum dot nanostructures". *New Journal of Physics* 7.1 (2005), p. 184. DOI: 10.1088/1367-2630/7/1/184.

[148] J. M. Blatt, K. Böer, and W. Brandt. "Bose-Einstein condensation of excitons". *Physical Review* 126.5 (1962), p. 1691. DOI: 10.1103/PhysRev.126.1691.

[149] J. Eisenstein and A. MacDonald. "Bose-Einstein condensation of excitons in bilayer electron systems". *Nature* 432.7018 (2004), pp. 691–694. DOI: 10.1038/nature03081.

[150] M. Alloing et al. "Evidence for a Bose-Einstein condensate of excitons". *Europhysics Letters* 107.1 (2014), p. 10012. DOI: 10.1209/0295-5075/107/10012.

[151] V. W. Scarola and S. D. Sarma. "Quantum phases of the extended Bose-Hubbard Hamiltonian: Possibility of a supersolid state of cold atoms in optical lattices". *Physical Review Letters* 95.3 (2005), p. 033003. DOI: 10.1103/PhysRevLett.95.033003.

[152] H.-Y. Wang et al. "Topological supersolidity of dipolar Fermi gases in a spin-dependent optical lattice". *Journal of Physics: Condensed Matter* 32.23 (2020), p. 235701. DOI: 10.1088/1361-648X/ab7871.

Acknowledgments

First and foremost I want to thank Henk, my supervisor, for giving me the best guidance I could've asked for in a project like this. Your physical intuition is unmatched, and your insights very often felt like a bright beam of light whenever I was completely stuck. I am proud of how much I've grown as a physicist during this last year, and can confidently say that you've had a great deal to do with it. I thoroughly enjoyed all our meetings and discussions and look forward to continue working with you in the years to come.

En segon lloc, moltes, moltíssimes gràcies a la meva família a Barcelona. Us trobo molt a faltar cada dia però em sento molt afortunat de poder comptar amb vosaltres per a qualsevol cosa. Papis i avis, em feu sentir molt feliç sempre que parlem i el suport i recolzament que em doneu sempre és una enorme ajuda per anar tirant endavant i fent el meu propi camí a la vida. I Fèlix, el meu 弟弟, estic molt orgullós de tu i sé que tens un futur brillant per davant. Persegueix sempre els teus somnis.

And last but not least, I would like to thank my friends. All of you have made these two years here in Utrecht something truly special and I wish you all the best in whatever you decide to do next. And for those who, just like me, decide to stay for a little more, I can't wait to share more funny moments, barbeques, and wonderful trips. I especially want to thank Gábor and Tami for being the most awesome flatmates someone can possibly find; may the silly jokes and adventures never end. Y por supuesto gracias a Melissa, a quien me siento increíblemente afortunado de haber conocido y compartido tantos momentos divertidos e inolvidables. Espero que sigamos haciendo el payaso y algún día encontremos los planetas.