

Quantum Hall Effect and Surface Criticality in 3D Topological Insulators

Inaugural-Dissertation

zur

Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Universität zu Köln

vorgelegt von

Matthias Sitte

aus

Olpe



Köln 2012

Berichterstatter: Prof. Dr. Achim Rosch
Prof. Dr. Hans Kroha

Tag der mündlichen Prüfung: 15. Oktober 2012

Abstract

Topological insulators are unique quantum states of matter. Although they behave like ordinary insulators in the bulk, states involving massless Dirac fermions at the surface of such materials are found to be *metallic*. Since their discovery in two-dimensional HgTe quantum wells, topological insulators have been very actively studied, both experimentally and theoretically.

In this thesis, we investigate how a magnetic field induces one-dimensional edge channels when the two-dimensional surface states of three-dimensional topological insulators become gapped. The Hall effect, which can be measured by contacting those edge channels, remains *integer*-quantized even when the topological θ -term in the bulk (and the associated surface Hall conductivities $\sigma_{xy}^{\text{surf}}$) are *not* quantized due to time-reversal symmetry breaking. We show that the quantization of the Hall conductivity $\sigma_{xy}^{\text{Hall}}$ arises as the θ -term changes by $\pm 2\pi n$ along a loop around n edge channels. Analytical calculations on a model based on strained HgTe are presented, which show how the interplay of orbital and Zeeman effects leads to quantum Hall transitions where channels get redistributed along the edges of the crystal. This network of edges, the existence of which we confirm by numerical tight-binding calculations, opens up new possibilities to investigate the coupling of edge channels.

In the last part of this thesis, we investigate whether long-ranged Coulomb interactions, controlled by the dimensionless coupling constant $\alpha = 1/(\hbar ev_F^{\text{surf}} \epsilon)$, can induce spontaneous symmetry-breaking on the surfaces of a three-dimensional topological insulator, thereby creating a gap in the metallic surface states. This would allow an anomalous quantum Hall effect without explicitly breaking time-reversal invariance, *i.e.*, without the application of an external magnetic field. We find that one prerequisite for observing this effect is to reduce the Fermi velocity v_F^{surf} of the surface Dirac fermions. However, we find that screening due to bulk metallic states renders the effective interaction strength α small instead of large and therefore prevents chiral symmetry breaking. We confirm this scenario by explicit numerical tight-binding calculations for various models in slab geometries, and by an analytical calculation of the corresponding polarization functions. We also derive topological criteria for the existence of flat surface bands and discuss under which conditions short-ranged Hubbard interactions may lead to an interaction-induced band gap in the surface states. In particular, we find a generic scenario for a surface band gap due to local interactions in the Fu-Kane-Mele model.

Kurzzusammenfassung

Topologische Isolatoren stellen eine neue Klasse von Materialien dar, die sich im Volumen wie gewöhnliche Isolatoren verhalten, auf ihrer Oberfläche aber metallische Zustände besitzen. Diese metallischen Randzustände lassen sich durch Dirac-Fermionen, d.h. durch masselose ultra-relativistische Teilchen mit linearer Energie-Impuls-Relation beschreiben und weisen außergewöhnliche Eigenschaften wie eine halbzahlige quantisierte Hall-Leitfähigkeit oder den sog. magneto-elektrischen Effekt auf. Seit ihrer Entdeckung in HgTe-Heterostrukturen vor wenigen Jahren sind topologische Isolatoren ein sehr aktives Forschungsgebiet.

In dieser Arbeit betrachten wir zunächst dreidimensionale topologische Isolatoren in Anwesenheit eines äußeren Magnetfeldes und untersuchen die Netzwerke eindimensionaler Randkanäle der zweidimensionalen Oberflächenzustände. Wir zeigen, dass die entsprechende Hall-Leitfähigkeit $\sigma_{xy}^{\text{Hall}}$ stets in Vielfachen von e^2/h quantisiert ist, wohingegen der sogenannte θ -Term und die verwandte Oberflächen-Leitfähigkeit $\sigma_{xy}^{\text{surf}}$ auf Grund gebrochener Zeitumkehrinvarianz keine Quantisierung mehr erfahren. Wir analysieren mittels Modellrechnungen den Einfluss des äußeren Magnetfeldes, das zugleich an Ladungs- und Spin-Freiheitsgrade der Dirac-Fermionen koppelt, auf die Randkanäle und zeigen, dass dies zu einer Verschiebung der Quanten-Hall-Übergänge auf den Oberflächen des topologischen Isolators führt. Des Weiteren gehen wir kurz auf mögliche Anwendungen derartiger Randkanalnetzwerke ein.

Im letzten Teil der vorliegenden Arbeit untersuchen wir den Effekt von langreichweitigen Coulomb-Wechselwirkungen, die durch chirale Symmetriebrechung eine Bandlücke in den ansonsten metallischen Oberflächenzuständen erzeugen kann. Wir diskutieren die Voraussetzungen für chirale Symmetriebrechung und zeigen, dass es ausreicht, die Fermi-Geschwindigkeit der Oberflächen-Dirac-Fermionen zu verringern, sofern die Dielektrizitätskonstante des topologischen Isolators endlich bleibt. Allerdings beobachten wir, dass eine Reduzierung der Fermi-Geschwindigkeit in den betrachteten Modellen stets mit einer Divergenz der Polarisationsfunktion (und damit der Dielektrizitätskonstanten) einhergeht, sodass langreichweitige Coulomb-Wechselwirkungen abgeschirmt werden, was wir durch explizite numerische und analytische Rechnungen überprüft haben. Zum Schluss diskutieren wir unter welchen Bedingungen generisch flache Bänder in topologischen Isolatoren auftreten können, und zeigen mögliche Wege auf, den Effekt der chiralen Symmetriebrechung durch lokale Hubbard-Wechselwirkungen zu realisieren, um beispielsweise den magneto-elektrischen Effekt auf der Oberfläche eines dreidimensionalen topologischen Isolators zu beobachten.

Acknowledgments

First of all, I am very much obliged to my advisor Prof. Dr. Achim Rosch for numerous fruitful discussions and ideas, for his encouragement, and especially for giving me the opportunity to write this thesis.

Second, I would like to thank Dr. Lars Fritz who accompanied this work from the beginning. Much of the insight into the problem emerged from illuminating discussions with him.

I would also like to thank Prof. Dr. Hans Kroha for being my BCGS mentor since 2008 and for co-refereeing this thesis.

I enjoyed working in the Institute of Theoretical Physics at the University of Cologne and I would like to thank my colleagues for creating an enjoyable working atmosphere. In particular, this work benefited from many discussions with Michael Becker, Stefan Buhardt, Dr. Karin Everschor-Sitte, Jan Gelhausen, Jonathan Lux, Dr. Tobias Meng, Dr. Eran Sela, and Mario Zacharias. I thank Dr. Andrew Mitchell for helping me creating the \LaTeX layout of the thesis. I appreciate the help of Michael Becker, Kasper Duivenvoorden, Dr. Karin Everschor-Sitte, Jan Gelhausen, Dr. Andrew Mitchell, and Sebastian Schmittner for proofreading parts of this thesis and for finding and eliminating spelling errors. I learned a lot from a number of present and former colleagues of the Institute of Theoretical Physics and would like to acknowledge their advice and comments: Dr. Fabrizio Anuso, PD Dr. Ralf Bulla, Dr. Rolf Helmes, Dr. Peter Jung, Dr. Ákos Rapp, Dr. David Rasch, Prof. Dr. Simon Trebst, and Prof. Dr. Matthias Vojta.

Furthermore, I am grateful for the help of Dr. Andreas Sindermann who took care of many problems regarding technical and administration issues.

I acknowledge funding from the Bonn-Cologne Graduate School of Physics and Astronomy (BCGS).

Finally, I would like to thank my parents, my brother and my sister for their loving support, and most importantly Karin who has been a great partner encouraging me over the last couple of years.

Contents

1	Introduction	1
2	Topological Quantum Matter	5
2.1	Classical Hall effect	6
2.2	Integer Quantum Hall effect	9
2.3	The Hall conductance as a topological invariant	18
2.4	Quantum Hall effect without Landau levels: Haldane model	24
2.5	Quantum spin Hall insulator	27
3	Electronic Band Structure of Topological Insulators	35
3.1	Tight-binding approach to the electronic band structure	36
3.2	Spin-orbit interactions	38
3.3	Electronic band structure of HgTe and CdTe	40
3.4	CdTe/HgTe/CdTe quantum well heterostructures	51
3.5	Strained HgTe as three-dimensional topological insulator	61
3.6	Construction of a minimal model for topological insulators	63
3.7	Summary	74
4	Hall Conductance Quantization and the θ -Term in Topological Insulators	77
4.1	Quantum Hall states in the minimal model for strained 3D HgTe	79
4.2	Numerical results for the quantum Hall effect in the minimal model	96
4.3	The magneto-electric effect and the topological θ -term	101
4.4	Summary and discussion	112
5	Surface Criticality in 3D Topological Insulators	113
5.1	Coulomb interactions and chiral symmetry breaking	115
5.2	The Kane-Mele model of 2D topological insulators	120
5.3	The Fu-Kane-Mele model on the diamond lattice	136
5.4	Flat surface bands in the minimal model for HgTe	158
5.5	Spontaneous gap generation due to local interactions	166
5.6	Discussion and outlook	171
A	Molecular Orbitals and Overlap Parameters	A-1
A.1	Overlap parameters between an s orbital and a p orbital	A-3
A.2	Overlap parameters between two p orbitals	A-3
A.3	Overlap parameters in diamond and zinc-blende lattices	A-4

B	Strain and Stress in Diamond and Zinc-Blende Crystals	B-1
B.1	Elastic strain and stress	B-1
B.2	Hydrostatic pressure	B-4
B.3	Uniaxial stress	B-4
B.4	Biaxial strain in the (001) plane	B-6
B.5	Strained 3D HgTe samples	B-7
C	Topological Aspects of the Su-Schrieffer-Heeger Model	C-1
C.1	The Su-Schrieffer-Heeger Hamiltonian	C-1
C.2	Topological origin of zero-energy bound states at domain walls	C-4
C.3	Role of next-nearest neighbor hopping breaking chiral symmetry	C-7
	Bibliography	i

CHAPTER 1

Introduction

In the last decade, physicists have begun to question one of the most basic classifications of materials as being either conductors or insulators. It was only rather recently understood that this distinction, based on the ability to conduct an electric current, is not as sharp as it was once thought to be. The essential subtlety is that all real materials are of finite size and hence have surfaces which can have very different properties from those of the bulk.

A huge theoretical and experimental research effort has now established that band insulators can actually be divided into two classes: the first one describes the “ordinary” band insulators, in which electrons fully occupy valence bands, while all available conduction states are well-separated by a large energy gap. Since there are no low-energy excitations, the system is trivially insulating. The second class, however, are the **topological insulators**, in which the energy gap between the occupied and empty bands is fundamentally modified due to spin-orbit interactions of relativistic origin. In topological insulators, the coupling of the spin and orbital angular momentum of an electron leads to an inversion of the band gap: states formerly lying above the gap are now below, and vice versa. Similar to the Möbius strip, this “twist” in the electronic band structure cannot be simply unwound, and this feature leads to the notion of **topology** in such insulators. Another consequence of this winded band structure is the appearance of metallic states at the surfaces of topological insulators which, in general, do not exist in the ordinary band insulators. Theoretically, those topological surface states have been predicted to show a linear energy-momentum relation similar to Dirac fermions, and an unusual spin texture due to a locking of spin and momentum. The experimental confirmation of those metallic surface states in HgTe quantum wells spurred a large activity in this field of research, serving as a platform for both fundamental physics and technological applications like spintronics or quantum computing.

From a historical point of view, the first and probably best-known example of a topological insulator is the **integer quantum Hall effect** which shows a quantized Hall conductivity in a perpendicular magnetic field¹. Haldane later realized that the quantum Hall effect may also occur in systems without an external magnetic field, and proposed a model for spinless fermions on the two-dimensional honeycomb lattice under the influence of a fictitious inhomogeneous magnetic field². By including the spin degree of freedom and making the magnetic field spin-dependent, Kane and Mele restored the time-reversal symmetry of the Haldane model, suggesting that the single-layer graphene sheet with intrinsic spin-orbit interactions indeed constitutes a new quantum phase of matter—the **quantum spin Hall insulator**^{3,4}. Soon after, the quantum spin Hall effect was in fact observed experimentally in HgTe quantum wells which

possess a larger spin-orbit coupling than the graphene system due to its heavier constituent elements^{5,6}.

In general, **symmetry-protected topological insulators** like the quantum spin Hall insulator can be described as quantum states of matter which are insulating in the bulk, but exhibit topologically protected edge or surface states at their boundaries. The surface states of a three-dimensional topological insulator are gapless and protected against arbitrary deformations and perturbations of the Hamiltonian as long as (i) time-reversal invariance as the most fundamental symmetry of the Hamiltonian is preserved and (ii) the bulk band gap is not closed. In three dimensions, the different phases of topological insulators are characterized by four topological \mathbb{Z}_2 invariants ν_i , of which ν_0 is the most important.

In the simplest case, a three-dimensional topological insulator can be constructed by stacking layers of the two-dimensional quantum spin Hall insulator on top of each other. The helical edge states of these layers hybridize and form surface states which can be localized by disorder, since time-reversal symmetry does not protect such states. This state of matter is a **weak topological insulator** which has $\nu_0 = 0$. On the other hand, $\nu_0 = 1$ describes a **strong topological insulator** which cannot be interpreted in terms of the 2D quantum spin Hall insulator. To be precise, in a strong topological insulator, there exists an odd number of Dirac points on the surface of the crystal. In the simplest case of a single Dirac point, the two-dimensional surface metal can be described in terms of a Dirac Hamiltonian with linear spectrum.

In this thesis, we focus on the three-dimensional topological insulator as a novel quantum state of matter. We begin in chapter 2 with a brief introduction to topological insulators and explain the topological order in the quantum Hall effect. We also give a short introduction to the Haldane model² and the quantum spin Hall insulator introduced by Kane and Mele^{3,4,7,8}. A feature common to all these models is the **bulk-boundary correspondence** which relates the topological structure of the electronic bulk band structure to the presence of gapless boundary modes. We discuss conceptually how to calculate both analytically and numerically the boundary states of topological insulators.

In chapter 3, we first describe the **electronic band structure** of mercury telluride (HgTe) and cadmium telluride (CdTe), and discuss the discovery of the quantum spin Hall insulator in two-dimensional CdTe/HgTe/CdTe quantum well heterostructures^{5,6}. We go on to discuss briefly the topological properties of bulk HgTe samples under applied biaxial strain, realizing an instance of a strong topological insulator in three dimensions^{9,10}. Finally, we develop a minimal model for topological insulators in three dimensions, based on the states close to the Fermi level in strained HgTe.

In chapter 4, we investigate the **Hall response** of a strong topological insulator and discuss the appearance of chiral edge states. We discuss the situation, where time-reversal invariance is broken in a strong topological insulator, for example by an applied external magnetic field, and show that a band gap opens for the otherwise gapless surface states. Those surface states are then expected to become quantum Hall states which exhibit a half-integer quantized Hall conductance $\sigma_{xy}^{\text{surf}} = \pm e^2/(2h)$ (Ref. 10). We also discuss the related phenomenon of the **magneto-electric effect** which is regarded as one of the characteristic features of strong topological insulators¹¹⁻¹³. Although the Hall conductivity of the two-dimensional surface metal is thus expected to be quantized in half integers, it is known from transport experiments that charge quantization leads to a quantized Hall conductance with $\sigma_{xy}^{\text{Hall}} = n(e^2/h)$ and *integer* n (Ref. 14). To resolve this apparent contradiction, we develop an effective field theory for the surface Dirac fermions and compute the Landau level sequence in the presence of an external magnetic field. The results are confirmed by explicit calculations of the band structure and the edge channels using exact diagonalization methods. Furthermore, we discuss the role of

the topological θ -term in the effective action of topological insulators which is related to the half-integer quantized conductance $\sigma_{xy}^{\text{surf}}$, and show how it can be reconciled with the integer-quantized Hall response $\sigma_{xy}^{\text{Hall}}$ from transport measurements. In particular, we show that the value of θ and thus $\sigma_{xy}^{\text{surf}}$ are *not* quantized in a system with broken time-reversal symmetry and broken inversion symmetry, but the Hall response $\sigma_{xy}^{\text{Hall}}$ as a global measurement indeed remains quantized—as follows from the quantization of the elementary charge.

Chapter 5 focuses on the question whether a symmetry-broken state could appear on the surfaces of a three-dimensional topological insulator due to an interaction-induced surface band gap. We investigate long-ranged Coulomb interactions which can, in principle, lead to a symmetry-broken state due to **chiral symmetry breaking**, where the Dirac fermions at the surface exhibit a spontaneously generated mass^{15–17}. Once such a gap is opened, one expects to see, for example, the anomalous quantum Hall effect or the surface magneto-electric effect associated with topological insulators^{11–13}.

We start by reviewing the mechanism of chiral symmetry breaking of Dirac fermions in two spatial dimensions in the presence of long-range Coulomb interactions and briefly discuss the current status regarding chiral symmetry-breaking in the context of two-dimensional graphene sheets. We then introduce the two-dimensional **Kane-Mele model** as an effective low-energy model of a topological insulator on the two-dimensional graphene lattice^{3,4}. In particular, finite-size graphene ribbons with zigzag edges host zero-energy edge states which form completely flat bands when spin-orbit interactions are absent, although those edge states are not observed for armchair edges. We show how these flat bands are related to the topology of a whole family of one-parameter Bloch Hamiltonians, and we give explicit criteria for the existence of those zero-energy boundary states for topological insulators in two and three dimensions¹⁸. The flat bands pave a way to achieve chiral symmetry breaking, and we discuss under which conditions the one-dimensional edge states in graphene ribbons spontaneously develop long-range ferromagnetic order in the presence of Coulomb interactions.

The results on the 2D graphene lattice can be generalized to strong topological insulators described by the **Fu-Kane-Mele model** on the three-dimensional diamond lattice^{7,8}. For certain surfaces of 2D slabs, we predict the existence of zero-energy states similar to the Kane-Mele model which form completely flat bands. We confirm the existence of those surface bands by explicit tight-binding calculations and discuss the role of spin-orbit interactions generated by second-nearest neighbor hopping. Concerning Coulomb interactions, the corresponding surface Dirac fermions are expected to show a similar phase diagram as the Dirac fermions in graphene, since the 2D surface states become more and more flat as the spin-orbit coupling parameter is decreased. Furthermore, we show by an explicit calculation of the polarization function to leading order in the spin-orbit coupling that the topological criterion for the existence of flat surface bands in that model also implies a sharp peak in the density of states. As a consequence, long-range Coulomb interactions are screened, and chiral symmetry breaking is ultimately avoided. Nevertheless, due to the vanishing kinetic energy of the Dirac fermions in the flat bands, even *local* interactions can lead to spontaneous gap generation, and we discuss a possible scenario for this effect in a mean-field approach.

Another class of 3D topological insulators is exemplified by the minimal model developed in chapter 3, wherein a similar situation for chiral symmetry breaking may occur. In that system, we also find surface bands which become more and more flat as we approach the bulk critical point. In contrast to the Fu-Kane-Mele model, the origin of the flat surface bands is similar to “spectral pressure” exerted by the bulk conduction and valence bands rather than a topological argument. However, in this model, a surface quantum phase transition to a gapped state is also hampered by the diverging polarization function, which is discussed by means of

an explicit calculation. However, due to the flatness of the surface bands also local interaction become relevant close to the bulk quantum critical point, and we find a generic situation for an interaction-induced surface band gap in the Fu-Kane-Mele model. On the other hand, in the minimal HgTe model we find a diverging penetration length of the surface states into the bulk which results in a vanishing effective interaction strength for the surface states. The calculation of the susceptibility matrix shows that ferromagnetism on the surfaces of the 3D topological insulator is the leading instability, but this occurs very closely to the magnetic bulk instability, so that fine-tuning of the model parameters is required for an observation of such a surface band gap. In the last part of this chapter, we put our results in the broader context of 3D topological insulators, and discuss the relevance of chiral symmetry breaking as a general mechanism involving surface states of topological insulators.

In Ref. 19, we published the construction of the minimal model based on the states in 3D strained HgTe (cf. chapter 3) and the results of chapter 4 concerning the quantum Hall effect and the non-quantized θ -term in topological insulators. The results of chapter 5 are currently being prepared for publication²⁰.

CHAPTER 2

Topological Quantum Matter

ne of the most important concepts in condensed matter physics is the classification of states of matter by the principle of **spontaneous symmetry breaking**. For example, a crystalline solid breaks translational symmetries even though the fundamental inter-atomic Coulomb interactions are translationally invariant, or a magnet breaks rotation symmetries, although the responsible spin-spin interactions are isotropic, while a superconductor breaks spontaneously the $U(1)$ gauge symmetry of the condensate. Each of those states of matter is described by a unique **order parameter** which assumes a non-vanishing expectation value only in the symmetry-broken or ordered phase. The effective field theory, usually referred to as **Landau-Ginzburg theory**, is determined by the general properties of the underlying system and the symmetries of some local order parameter. For a long time it was believed that this theory provides a universal description of all quantum states of matter.

In 1980, however, von Klitzing discovered experimentally a new quantum state of matter—the so-called **integer quantum Hall effect**—which cannot be described in terms of a Landau-Ginzburg theory. In a nutshell, the basic observation was that the longitudinal conductance σ_{xx} of a two-dimensional electron gas subjected to a perpendicular magnetic field almost vanishes, while the transversal conductance or **Hall conductance** is an integer multiple of the conductance quantum $G_0 = e^2/h$ (Ref. 1). This quantization turned out to be independent of microscopic details such as the type of semiconducting material used, the quality of the sample, or the precise magnitude of the magnetic field. “For the discovery of the quantized Hall effect” the Nobel Prize in Physics was awarded to von Klitzing in 1985²¹. Actual measurements of the Hall conductance show that e^2/h can be measured with a precision of about 10^{-9} or better, which is a manifestation of charge quantization or, in other words, gauge invariance²². The quantum Hall effect also allows for a redefinition of the international standard for resistance in terms of the electron charge $-e$ and Planck’s constant h alone, based on the resistance quantum given by the von Klitzing constant $R_K = h/e^2 = 25812.807557(18) \Omega$ (Ref. 23).

Two years after the discovery of the integer quantum Hall effect Tsui, Störmer, and Gosard discovered in high-quality samples with very low disorder that the Hall conductance σ_{xy} also exhibits a fractional quantization^{24,25}. Strong Coulomb interactions and correlations between the electrons are indeed responsible for the so-called **fractional quantum Hall effect**. The quasi-particles condense into quantum states whose excitations show fractional quantum numbers, including fractional charge and fractional (anyonic) statistics. The Nobel Prize in Physics 1998 was awarded jointly to Laughlin, Störmer and Tsui “for their discovery of a new form of quantum fluid with fractionally charged excitations”²⁶.

Historically, the **quantum Hall effect** is the first example of topological quantum matter

which, in general, is characterized by non-local properties instead of a Ginzburg-Landau theory with a local order parameter. In this chapter, we begin with a brief review of the classical and quantum Hall effect, discussing in particular the appearance of chiral **edge states** in the quantum Hall state. After that we review the approach by Thouless, Kohmoto, Nightingale, and den Nijs describing the Hall conductance as a **topological invariant**¹⁴. Although there is also an alternative approach due to Laughlin which is based on the gauge invariance of the response of a system to a change in the external vector potential²², we do not consider Laughlin's argument, but rather focus on the approach by Thouless *et al.* as it demonstrates explicitly the concepts of topology used in this thesis.

Another important step is to consider the so-called **Haldane model** which, in the original work, borrowed the idea of the parity anomaly in quantum electrodynamics to construct a quantum Hall state on the 2D honeycomb lattice². In that model, the role of the external magnetic field is taken over by an intrinsic magnetic field generated through second-nearest neighbor interactions, and it turns out that this toy model shows a quantization of the Hall conductance similar to the quantum Hall effect. In both systems, fundamental symmetries like time-reversal invariance, translation invariance, charge-conjugation (particle-hole) symmetry, or chiral symmetry are broken by the applied magnetic field or disorder in the samples. In that sense, the quantum Hall effect and related system represent a certain class of **intrinsic topological insulators**, whose properties will be studied later.

A second class is given by the so-called **symmetry-protected topological insulators** which have been studied extensively in the last decade. As was shown by Kane and Mele in 2005, one can construct a certain model on the graphene lattice with non-trivial topological properties by adding up two time-reversed copies of the Haldane model, thereby restoring time-reversal symmetry^{3,4}. In contrast to the intrinsic topological insulators, the resulting model shows a quantization of spin currents instead of charge currents which has led to the notion of the **quantum spin Hall insulator**. The existence of the quantum spin Hall insulator as a two-dimensional **topological insulator** was first predicted by Bernevig, Hughes, and Zhang in 2006⁵, and was soon after realized by König *et al.* in CdTe/HgTe/CdTe quantum well heterostructures^{6,27}. The experimental discovery of the two-dimensional topological insulators has spurred a huge interest into topological quantum matter, both experimentally and theoretically.

The generalization of the TKNN invariant to the quantum spin Hall state by Kane and Mele has also shown that all time-reversal invariant topological insulators fall into two distinct topological classes (without ground state degeneracies) which are classified by a \mathbb{Z}_2 **topological invariant**^{3,4,7,8}. Although this approach is based on the **topological aspects of band theory** and is thus only valid for non-interacting systems, it has become an important tool in the discovery of new topological materials¹³. In the last part of this chapter, we briefly review the topological band theory and the \mathbb{Z}_2 classification of symmetry-protected topological insulators in two and three dimensions as introduced by Fu, Kane, and Mele^{3,4,7,8,28}.

Finally, note that this is an introductory chapter, and thus it is intended to provide a broad overview of the analytical and numerical concepts used in this thesis. As such, this chapter does not cover more recent developments like, for example, dynamical effects.

2.1 Classical Hall effect

Let us first discuss the classical Hall effect. In a two-dimensional electron gas in the xy plane, the application of an electric field $\mathbf{E} = E_x \mathbf{e}_x$ leads to a current density j_x flowing in x direction. However, in a perpendicular magnetic field $\mathbf{B} = B \mathbf{e}_z$ (see Fig. 2.1), the Lorentz force

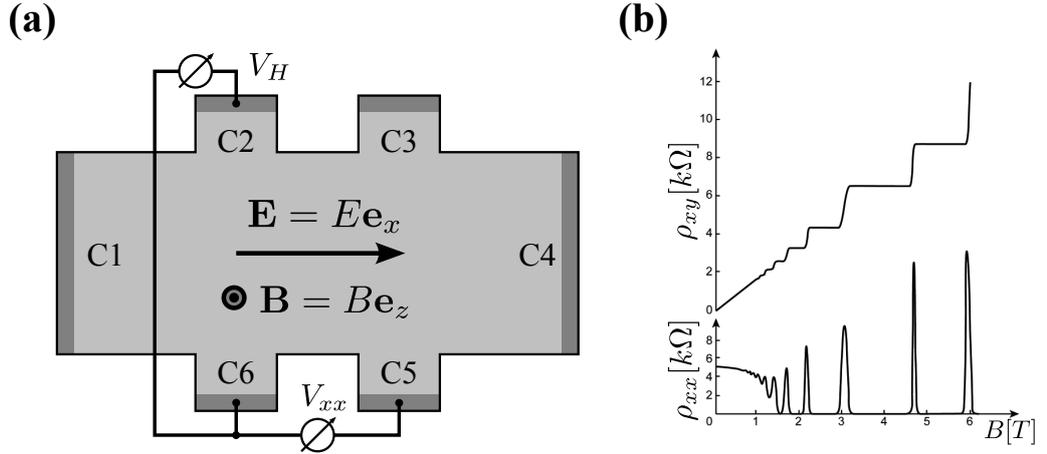


Figure 2.1. (a) Schematic view of a typical four-terminal setup for Hall measurements. An electron gas confined to two dimensions is subjected to an electric field $\mathbf{E} = E\mathbf{e}_x$ within the plane and a perpendicular magnetic field $\mathbf{B} = B\mathbf{e}_z$. The voltage drop V_{xx} in longitudinal direction is measured between contacts C_5 and C_6 , while the Hall voltage V_H is measured between contacts C_2 and C_6 . (b) Sketch of the Hall resistance ρ_{xy} and of the longitudinal resistance ρ_{xx} as function of the applied magnetic field B . In the classical or low-field Hall regime, the transverse or Hall resistance ρ_{xy} is proportional to the magnetic field, while for larger magnetic field strengths ρ_{xy} exhibits plateaus. Figure after Ref. 29.

$\mathbf{F}_L = (q/m)(\mathbf{p} \times \mathbf{B})$ deflects the electrons with charge $q = -e$ and mass m drifting in the x direction with momentum \mathbf{p} into the y direction. As a result of the accumulating charge at the surfaces, an electric field E_y builds up in y direction so that in equilibrium this electric field will balance the Lorentz force, and the current will flow again only in the x direction. In that system, two quantities are of particular interest: First, the so-called longitudinal resistance or **magneto-resistance** is given by the ratio of the external field E_x and the current density j_x , while the transversal resistance or **Hall resistance** relates j_x to the transverse electric field E_y :

$$\rho_{xx} \equiv \frac{E_x}{j_x}, \quad \rho_{xy} \equiv \frac{E_y}{j_x}. \quad (2.1)$$

To gain a qualitative understanding of the classical Hall effect, let us consider the Drude theory of metals, describing the diffusive transport of non-interacting electrons in solids. The electrons with mass m and charge $q = -e$ are subject to the classical equations of motion,

$$\dot{\mathbf{p}} = q \left(\mathbf{E} + \frac{\mathbf{p}}{m} \times \mathbf{B} \right) - \tau^{-1} \mathbf{p}. \quad (2.2)$$

The first term is the usual Lorentz force acting on charged particles, while the last term describes the momentum relaxation due to scattering of the charge carriers from impurities with a characteristic **relaxation time** τ . In the steady state, the macroscopic transport is independent of time, and therefore we find the following set of equations relating the electric fields to the current densities:

$$0 = qE_x - \omega_c p_y - \tau^{-1} p_x, \quad (2.3a)$$

$$0 = qE_y + \omega_c p_x - \tau^{-1} p_y. \quad (2.3b)$$

Here, we have defined the so-called **cyclotron frequency** by

$$\omega_c \equiv \frac{qB}{m}. \quad (2.4)$$

This quantity characterizes the cyclotron motion of a charged particle in a perpendicular magnetic field due to the Lorentz force. By multiplying these equations by $n_q q \tau / m$, where n_q is the density of charged particles, we find that

$$\sigma_0 E_x = -\omega_c \tau j_y + j_x, \quad (2.5a)$$

$$\sigma_0 E_y = \omega_c \tau j_x + j_y. \quad (2.5b)$$

Here, we have introduced the current density $\mathbf{j} \equiv q n_q \mathbf{v} = q n_q \mathbf{p} / m$, and the **Drude conductivity** is given by

$$\sigma_0 \equiv \frac{n_q q^2 \tau}{m}. \quad (2.6)$$

Obviously, the perpendicular magnetic field \mathbf{B} leads to non-zero off-diagonal matrix elements of the resistivity tensor ρ which relates the electric field $\mathbf{E} = \rho \mathbf{j}$ to the current density \mathbf{j} :

$$\rho = \frac{1}{\sigma_0} \begin{bmatrix} 1 & -\omega_c \tau \\ \omega_c \tau & 1 \end{bmatrix}. \quad (2.7)$$

In the static limit, the electric field E_y is determined by the condition that there is no transverse current in y direction, $j_y = 0$. Solving Eq. (2.5b) for $j_y = 0$ we find that

$$E_y = \frac{\omega_c \tau}{\sigma_0} j_x = \frac{B}{n_q q} j_x. \quad (2.8)$$

One can now easily read off the Hall resistance as

$$\rho_{xy} = -\rho_{yx} = -\frac{E_y}{j_x} = -\frac{\omega_c \tau}{\sigma_0} = \frac{B}{(-q)n_q}. \quad (2.9)$$

Here, some aspects are worth noting: First, in the semi-classical limit, the Hall resistance ρ_{xy} of a one-band system is completely independent of microscopic details like, for example, the relaxation time τ which strongly depends on the nature of inter-particle interactions or the type of material used. Second, Eq. (2.9) shows that the Hall resistance ρ_{xy} grows linearly with the applied magnetic field B [see Fig. 2.1 (b)], while the slope of the Hall resistance $\rho_{xy}(B)$ provides access to the nature of the charge carriers and their density n_q . Note that $\rho_{xy}(B)$ has a positive slope for particles with negative charge like electrons.

Finally, the conductivity tensor σ defined by Ohm's law $\mathbf{j} = \sigma \mathbf{E}$ follows from the resistivity tensor ρ by straightforward matrix inversion:

$$\sigma = \frac{\sigma_0}{1 + \omega_c^2 \tau^2} \begin{bmatrix} 1 & -\omega_c \tau \\ \omega_c \tau & 1 \end{bmatrix}. \quad (2.10)$$

If we consider a clean system subjected to a magnetic field, *i.e.*, we take the limit $\omega_c \tau \rightarrow \infty$ with finite cyclotron frequency ω_c , the resistivity and conductivity tensors are given by

$$\rho = \rho_{xy} \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \sigma = \frac{1}{\rho_{xy}} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}. \quad (2.11)$$

In that limit, the longitudinal components of the resistivity tensor ρ vanish. However, also the longitudinal components of the conductivity tensor σ vanish as well despite the fact that we consider a clean system here, where one would naively expect perfect conductance across the sample. The resolution of this seemingly counterintuitive result is that in the presence of an external magnetic field the charged particles move on closed orbits, leading to finite off-diagonal

elements in the resistivity tensor with $\rho_{xy} \neq 0$ [cf. Eq. (2.9)]. As a consequence, $\sigma_{xy} = -1/\rho_{xy}$ is finite as well.

Finally, note that the limits $\omega_c \rightarrow 0$ and $\tau \rightarrow \infty$ do not commute. If we had taken the limit $\omega_c \rightarrow 0$ first, the conductivity tensor simply becomes a diagonal matrix proportional to the Drude conductivity σ_0 , where the scattering time enters as $\sigma_0 = n_q q^2 \tau / m$, and we recover the usual conductivity of a normal metal in the absence of a magnetic field.

2.2 Integer Quantum Hall effect

The semi-classical linear behavior of the Hall resistivity as function of the applied magnetic field has been observed experimentally in two-dimensional electron gases at weak magnetic fields. By the end of the 1970s, the fabrication techniques of heterostructure devices had advanced so far that one could produce two-dimensional electron gases with high mobilities at the interfaces of semiconducting materials. In 1980, von Klitzing discovered that in such an electron gas the Hall resistance ρ_{xy} develops plateau structures when it is subjected to strong magnetic fields (of about 15 – 20 T) and cooled to very low temperatures (of about 1.5 K), thus invalidating the classical picture of the Hall effect¹. This striking phenomenon became known as the **integer quantum Hall effect** and was recognized with the Nobel prize in Physics in 1985²¹. Fig. 2.2 shows the transport data for a real device in the quantum Hall regime. Instead of a linear Hall resistivity [cf. Eq. (2.9)] we observe a series of so-called **Hall plateaus** in the Hall resistance ρ_{xy} or equivalently in the Hall conductance σ_{xy} which are given by

$$\rho_{xy} = \frac{1}{\nu} \frac{h}{e^2} \quad \text{or} \quad \sigma_{xy} = \nu \frac{e^2}{h}. \quad (2.12)$$

Experimentally, the quantization of σ_{xy} with integer ν turns out to be independent of microscopic details such as the type of semiconducting material used, the quality of the sample, and the precise magnitude of the magnetic field. At each of these plateaus, the longitudinal resistance ρ_{xx} vanishes by many orders of magnitude. As suggested by the name, the quantum Hall effect requires inherently a quantum mechanical description of the electrons in two dimensions. In this section, we first discuss the Hall effect of a two-dimensional electron gas from a simple quantum mechanical point of view and derive the integer quantization of the Hall conductance σ_{xy} , following the two reviews by Girvin³⁰ and Goerbig³¹. Later, we review the topological argument of Thouless, Kohmoto, Nightingale, and den Nijs for the quantization of σ_{xy} (Ref. 14).

2.2.1 Two-dimensional electrons in a magnetic field

To gain a basic understanding of the quantum Hall physics, we consider a free electron gas whose movement is confined to the two-dimensional xy plane in an applied perpendicular magnetic field $\mathbf{B} = B\mathbf{e}_z$. We further assume that the electron spins are polarized, *i.e.*, we ignore the electronic spin degree of freedom in the following. Recall from elementary quantum mechanics that a single electron with charge $-e$ in a uniform magnetic field is described by the following Hamiltonian (throughout this thesis we choose units such that $c = 1$):

$$H = \frac{1}{2m} (\mathbf{p} + e\mathbf{A}(\mathbf{r}))^2 = \frac{\mathbf{\Pi}^2}{2m}. \quad (2.13)$$

Here, we have replaced the usual momentum operator \mathbf{p} by the **canonical momentum** operator $\mathbf{\Pi} \equiv \mathbf{p} + e\mathbf{A}(\mathbf{r})$, with $\mathbf{A}(\mathbf{r})$ the vector potential generating the magnetic field, $\mathbf{B}(\mathbf{r}) =$

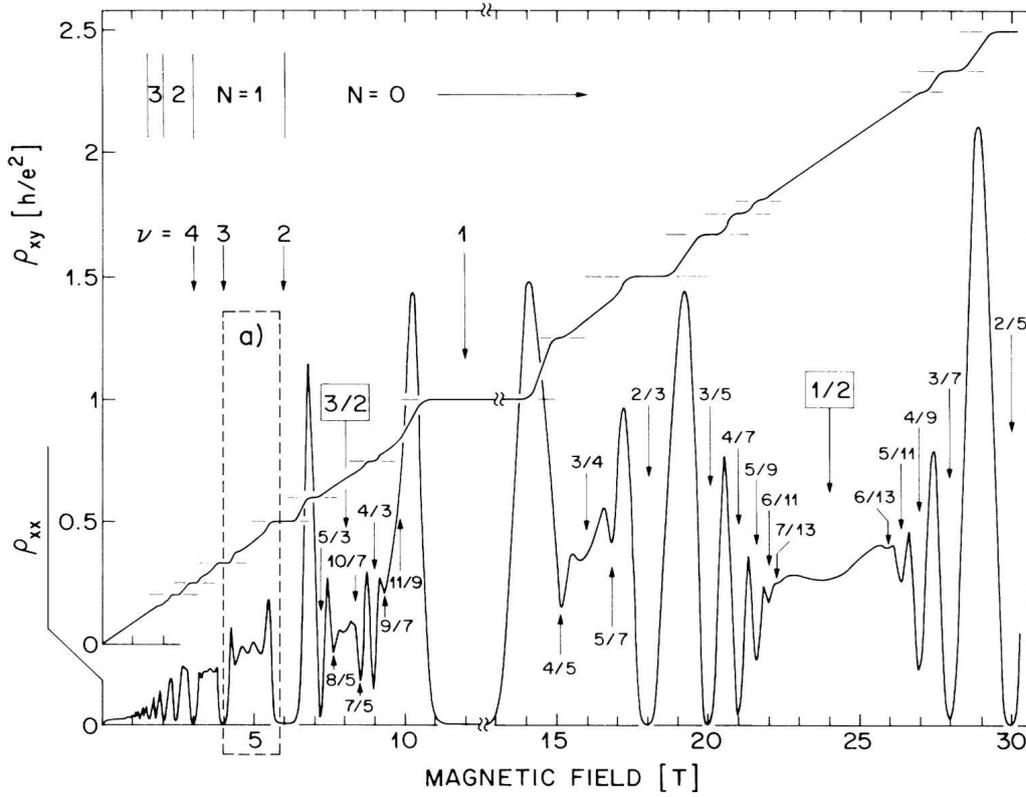


Figure 2.2. Integer and fractional quantum Hall transport data showing the plateau regions in the Hall resistance ρ_{xy} and associated minima in the longitudinal resistance ρ_{xx} for a single-interface GaAs/AlGaAs heterostructure grown by molecular beam epitaxy. For magnetic fields below 12 T the temperature used was about 150 mK, while the high-field measurements were performed at 85 mK. The numbers ν and N indicate the filling factors ν and the Landau levels at which various features occur, as discussed in the original publication by Willet *et al.* (see Ref. 25).

$\text{curl } \mathbf{A}(\mathbf{r})$. Note that neither $\mathbf{A}(\mathbf{r})$ nor \mathbf{p} are gauge invariant, but $\mathbf{\Pi}$ actually is. A straightforward calculation then shows that the components of the canonical momentum $\mathbf{\Pi}$ no longer commute, but instead obey the following commutation relations:

$$[\Pi_x, \Pi_x] = [\Pi_y, \Pi_y] = 0, \quad [\Pi_x, \Pi_y] = -\frac{\hbar^2}{l_B^2}, \quad (2.14)$$

where we have introduced the **magnetic length** as

$$l_B \equiv \sqrt{\frac{\hbar}{eB}}. \quad (2.15)$$

l_B is the fundamental length scale in the presence of a magnetic field. The physical interpretation of this length scale is that the area $2\pi l_B^2$ contains one **quantum of magnetic flux**, $\Phi_0 = h/e$. In other words, the density of magnetic flux is given by

$$B = \frac{\Phi_0}{2\pi l_B^2}. \quad (2.16)$$

As a consequence of the non-vanishing commutator $[\Pi_x, \Pi_y] = -\hbar^2/l_B^2$ we cannot find common eigenfunctions to Π_x and Π_y at the same time, in contrast to the zero-field limit, where

the momenta p_x and p_y commute. In that sense, even a uniform magnetic field “breaks” the translational invariance of the two-dimensional electron gas. Note that any gauge of the vector potential breaks translational invariance in at least one direction, and thus the momentum in that directions is no longer a good quantum number to describe the quantum states of the two-dimensional electron gas.

This fact becomes apparent for certain choices of the vector potential. The so-called **symmetric gauge** and the **Landau gauge** are two convenient and frequently used choices for the vector potential $\mathbf{A}(\mathbf{r})$. For a homogeneous magnetic field pointing in z direction, *i.e.*, $\mathbf{B}(\mathbf{r}) = B\mathbf{e}_z$, those vector potentials are given by

$$\mathbf{A}_{\text{symm}}(\mathbf{r}) = B(x\mathbf{e}_y - y\mathbf{e}_x)/2, \quad (2.17a)$$

$$\mathbf{A}_{\text{Landau}}(\mathbf{r}) = -By\mathbf{e}_x. \quad (2.17b)$$

The Landau gauge is particularly useful when it comes to band structure calculations, because in this gauge the vector potential $\mathbf{A}(\mathbf{r})$ points in the x direction, but varies only with the y position. Hence, the system still possesses translational invariance in the x direction, and the corresponding momentum k_x remains a good quantum number which can be used to calculate the band structure of the quantum Hall system (see below).

To solve the stationary Schrödinger equation for the Hamiltonian (2.13) we consider the Landau gauge, $\mathbf{A} = -By\mathbf{e}_x$, and rewrite the Hamiltonian in terms of the kinetic momenta p_x and p_y as

$$H = \frac{1}{2m} [(p_x - eBy)^2 + p_y^2]. \quad (2.18)$$

To take advantage of the translational symmetry in x direction let us consider a plane-wave ansatz for the wave function of the form

$$\psi_k(x, y) = e^{ikx} \phi_k(y) \quad (2.19)$$

which is an eigenstate of the p_x operator with eigenvalue $\hbar k$. Substituting this ansatz into the Hamiltonian (2.13) and multiplying by $\exp(-i\hbar k)$ from the left we obtain the effective one-dimensional Schrödinger equation

$$h_k \phi_k(y) = \mathcal{E}_k \phi_k(y), \quad (2.20)$$

where h_k is the Hamiltonian of a one-dimensional displaced harmonic oscillator:

$$h_k = \frac{p_y^2}{2m} + \frac{m\omega_c^2}{2} (y + kl_B^2)^2. \quad (2.21)$$

Here, the frequency of the harmonic oscillator is determined by the cyclotron frequency ω_c [cf. Eq. (2.4)], and its central position $Y_k = -kl_B^2$ is proportional to the momentum k . From quantum mechanics we recall that the eigenvalues of the above Hamiltonian are simply given by

$$\mathcal{E}_{n,k} = \hbar\omega_c \left(n + \frac{1}{2} \right). \quad (2.22)$$

Thus, for each plane wave with momentum $\hbar k$ in the x direction there will be an entire set of energy eigenvalues labelled by an integer n , and those energy levels are commonly known as the **Landau levels** of the two-dimensional electron gas. They form completely flat, *i.e.*, dispersionless “bands” as they are independent of the momentum k , leading to the electronic band structure of a quantum Hall sample shown schematically in Fig. 2.3 (a). Note that generically

the chemical potential lies in between the Landau levels, rendering the quantum Hall sample a true bulk insulator. When the chemical potential of the system is fine-tuned to coincide with one of the Landau levels, the system becomes metallic, leading to sharp steps in the Hall conductance σ_{xy} . Those steps are, however, slightly broadened in a real sample due to disorder effects in the bulk of the sample. As a consequence, in general one would not expect to observe any finite conductance in the system, *i.e.*, naively $\sigma_{xx} = \sigma_{xy} = 0$. Nonetheless, the experimentally observed quantization of the Hall conductance $\sigma_{xy} = \nu(e^2/h)$ is not carried by the bulk states of the sample, but rather stems from an integer number of edge states present at the sample boundaries, as we show below.

A second observation from Eq. (2.22) is that the Landau levels are hugely degenerate as they are completely independent of the momentum $\hbar k$ in the x direction. To determine the degeneracy of the Landau levels, let us consider a rectangular sample with dimensions $L_x \times L_y$ with $L_x, L_y \gg l_B$ and periodic boundary conditions in the x direction. As mentioned above, it is impossible to construct common eigenfunctions of Π_x and Π_y , and therefore it is impossible to also have periodic boundary conditions in the y direction. The eigenfunctions of the harmonic oscillator (2.21) are strongly peaked at their center $Y_k = -kl_B^2$,

$$\psi_{k,n}(\mathbf{r}) \propto e^{ikx} H_n(y - Y_k) \exp\left[-\frac{1}{2l_B^2} (y - Y_k)^2\right], \quad (2.23)$$

where $H_n(x)$ is the n th Hermite polynomial, and the translational invariance in x direction is taken into account by the plane-wave ansatz $\exp(ikx)$. Note that the width of the harmonic oscillator wave functions in the n th Landau level is of the order $\sqrt{n} l_B$ which is microscopic compared to the system size, but the spacing between the centers of the wave functions is much smaller, $2\pi l_B^2/L_x$ because $L_x \gg l_B$. Consequently, the values of the wave vector k for which the basis state is inside the sample run from $k = 0$ to $k = L_x/l_B^2$. The total number of states in each Landau level is then

$$N = \frac{L_y}{2\pi} \int_0^{L_x/l_B^2} dk = \frac{L_x L_y}{2\pi l_B^2} = \frac{\Phi}{\Phi_0} \equiv N_\Phi. \quad (2.24)$$

Here,

$$N_\Phi \equiv \frac{\Phi}{\Phi_0} = \frac{BL_x L_y}{\Phi_0} \quad (2.25)$$

is the number of flux quanta penetrating the sample. Importantly, there is one state per Landau level per flux quantum. Since $N_\Phi \gg 1$ we find that the degeneracy of each Landau level is huge. The reason for this is that the spacing $\Delta k = 2\pi/L_x$ between the wave vectors k allowed by periodic boundary conditions decreases as function of sample size L_x , while the range of allowed wave vectors $[0, L_x/l_B^2]$ increases.

Finally, note that the above results for the Landau levels (2.22) are independent of the particular gauge of the vector potential $\mathbf{A}(\mathbf{r})$, whereas the corresponding eigenfunctions take the particular form of Eq. (2.23) only in the Landau gauge. To solve the Hamiltonian (2.13) for a generic vector potential it is convenient to introduce a pair of raising and lowering operators a, a^\dagger similar to the quantum mechanical treatment of the harmonic oscillator. In the case of the two-dimensional electron gas in a magnetic field, those raising and lowering operators are given by

$$a \equiv \frac{l_B}{\sqrt{2\hbar}} (\Pi_x - i\Pi_y), \quad a^\dagger \equiv \frac{l_B}{\sqrt{2\hbar}} (\Pi_x + i\Pi_y), \quad (2.26)$$

where the prefactors ensure a proper normalization with respect to the commutator:

$$[a, a^\dagger] = 1. \quad (2.27)$$

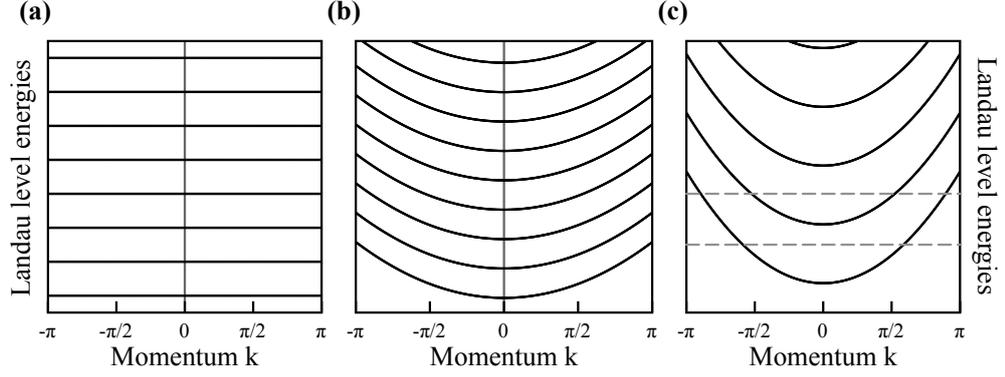


Figure 2.3. (a) Schematic picture of the Landau levels for the homogeneous system in the absence of disorder. In the Landau gauge, one can plot the Landau levels as function of the momentum k , where they form “bands” which are completely flat, *i.e.*, dispersionless. (b) In the presence of a harmonic confinement potential in y direction, $V_{\text{conf}}(y) = m\omega_{\text{conf}}(y^2/2)$, those bands acquire a finite curvature as function of the momentum k . (c) Zoom into panel (b). In contrast to panel (a), for any chemical potential there is an equal number of dispersing channels with positive and negative group velocity traversing the sample at opposite edges of the sample.

In terms of the gauge-invariant momentum $\mathbf{\Pi}$, the Hamiltonian (2.13) is given by

$$H = \frac{\mathbf{\Pi}^2}{2m} = \frac{1}{2m} (\Pi_x^2 + \Pi_y^2). \quad (2.28)$$

Substituting the ladder operators a and a^\dagger defined in Eq. (2.26) this Hamiltonian may be rewritten in the usual form of harmonic oscillators:

$$H = \hbar\omega_c \left(a^\dagger a + \frac{1}{2} \right), \quad (2.29)$$

where we have used the relation $\omega_c = \hbar/(ml_B^2)$ between the cyclotron frequency ω_c [Eq. (2.4)] and the magnetic length l_B [Eq. (2.15)]. The eigenvalues and eigenstates of the above Hamiltonian are those of the number operator $\hat{n} = a^\dagger a$. Thus, the energy levels of the two-dimensional electron gas are given by

$$\mathcal{E}_{n,n'} = \hbar\omega_c \left(n + \frac{1}{2} \right), \quad (2.30)$$

and we obtain the same result as before. Although the original Hamiltonian for the two-dimensional electron gas is formulated in terms of two pairs of conjugate operators, *i.e.*, $\{x, p_x\}$ and $\{y, p_y\}$, and one expects in general the eigenvalues to depend on two independent quantum numbers n, n' , we find that the Landau level sequence depends only on n . As a consequence, they are highly degenerate, as shown above. By choosing a suitable gauge like the Landau gauge, $\mathbf{A}(\mathbf{r}) = -By\mathbf{e}_x$, we have seen that the second quantum number n' can be interpreted as, for example, the conserved momentum k along the x direction. Since this allows us to plot meaningful band structures as function of k , in the following we always consider the Landau gauge whenever a uniform magnetic field is applied.

2.2.2 The appearance of boundary states

So far we have considered the homogeneous quantum Hall sample which turns out to be a true bulk insulator, unless the chemical potential is fine-tuned to coincide with one of the Landau

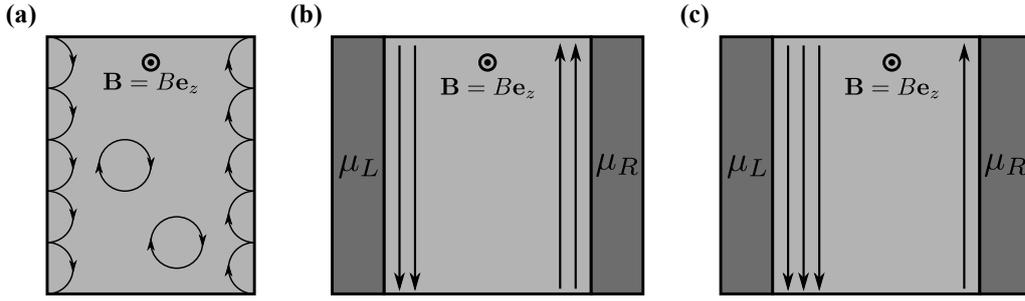


Figure 2.4. (a) Illustration of the skipping orbits: In a finite-size system, the boundaries prevent electrons from completing a full revolution, leading to chiral states located at the boundaries of the sample. (b) For larger magnetic fields, higher Landau levels are populated, and electrons in those Landau levels also form chiral edge states. For example, we can observe two chiral edge states at each boundary of the sample if two Landau levels are activated ($\mu_L = \mu_R$). Moreover, by contacting the side surfaces of a quantum Hall bar we may perform conductance measurements probing the number of edge states. (c) If the electrochemical potentials lie between different Landau levels, *i.e.*, $\mu_L \neq \mu_R$, the number of edge channels differs, and there is a net charge transport across the sample.

levels (2.22). However, the experimental observation of the quantization of the Hall conductance, $\sigma_{xy} = \nu(e^2/h)$ with integer ν , contradicts this seemingly innocuous argument. The solution to this paradox can be seen by explicitly considering the edges of the quantum Hall sample.

Let us first discuss a classical picture which, for apparent reasons, is often referred to as the **skipping orbit** interpretation. An electron subjected to a magnetic field will move on closed orbits within the plane normal to the direction of the applied magnetic field due to the Lorentz force. However, close to the boundaries of the sample the electron cannot complete the revolution, but instead “bounces off” of the sample boundary. This leads to a uni-directional motion of the electrons in one direction at the edge, while electrons at the opposite edge move in the opposite direction. Fig. 2.4 (a) shows a schematic illustration of those **chiral edge states** occurring at the boundaries of a quantum Hall sample.

We can now ask how this classical picture has to be modified when we take the quantum nature of the electrons into account. To show the appearance of edge channels in the quantum mechanical version we have to introduce the notion of “boundaries” into the model Hamiltonian. For simplicity, we consider a system with periodic boundary conditions in the x direction and add a harmonic confinement potential

$$V_{\text{conf}}(y) = \frac{1}{2} m\omega_{\text{conf}}^2 y^2 \quad (2.31)$$

in the y direction which is assumed to vary slowly of the scale of the magnetic length, *i.e.*, $\omega_{\text{conf}} \ll \omega_c$. This is a very crude approximation of the boundaries of a real quantum Hall sample, but allows for an analytic solution with respect to the Landau levels and the emerging chiral edge states. The single-particle Hamiltonian for a two-dimensional electron gas in the presence of the confinement potential $V_{\text{conf}}(y)$ is given by

$$H = \frac{\mathbf{\Pi}^2}{2m} + V(y) = \frac{1}{2m} [(p_x - eBy)^2 + p_y^2] + \frac{1}{2} m\omega_{\text{conf}}^2 y^2, \quad (2.32)$$

where we have employed the Landau gauge $\mathbf{A} = -By\mathbf{e}_x$ for the perpendicular magnetic field $\mathbf{B} = B\mathbf{e}_z$. As before, the translational symmetry of the system in x direction allows us to

choose eigenfunctions of the Hamiltonian which take the following form:

$$\psi_k(x, y) = e^{ikx} \phi_k(y). \quad (2.33)$$

This particular ansatz for a wave function leads to the one-dimensional Schrödinger equation $h_k \phi_k(y) = \mathcal{E}_k \phi_k(y)$ for $\phi_k(y)$ with the following single-particle Hamiltonian:

$$h_k = \frac{p_y^2}{2m} + \frac{1}{2} m \omega_c^2 (y - kl_B^2)^2 + \frac{1}{2} m \omega_{\text{conf}}^2 y^2. \quad (2.34)$$

Completing the square we see that this is again the Hamiltonian of a displaced harmonic oscillator:

$$h_k = \frac{p_y^2}{2m} + \frac{m \bar{\omega}_c^2}{2} (y + \bar{Y}_k)^2 - \frac{\omega_{\text{conf}}^2}{\omega_c^2 - \omega_{\text{conf}}^2} \frac{\hbar^2 k^2}{2m}. \quad (2.35)$$

Apparently, the displacement is different compared to the homogeneous case, *i.e.*, the harmonic oscillator is now centered at the position

$$\bar{Y}_k = -kl_B^2 \frac{\omega_c^2}{\omega_c^2 - \omega_{\text{conf}}^2}, \quad (2.36)$$

and the frequency of the harmonic oscillator is detuned from the cyclotron frequency:

$$\bar{\omega}_c \equiv \sqrt{\omega_c^2 - \omega_{\text{conf}}^2}. \quad (2.37)$$

Importantly, in the presence of the confinement potential $V_{\text{conf}}(y)$, the set of energy eigenvalues does not only depend on the integer n labeling the Landau levels, but also shows dispersive behavior as function of the momentum $\hbar k$:

$$\mathcal{E}_{n,k} = \hbar \bar{\omega}_c \left(n + \frac{1}{2} \right) + \frac{\hbar^2 k^2}{2m'}, \quad (2.38)$$

where we have introduced the effective mass $m' \equiv m(\omega_c^2 - \omega_{\text{conf}}^2)/\omega_{\text{conf}}^2$. Fig. 2.3 (b) shows the Landau levels as function of the momentum $\hbar k$ in the presence of the confinement potential $V_{\text{conf}}(y)$. By turning off the confinement potential, $\omega_{\text{conf}} \rightarrow 0$, the effective mass diverges, $m' \rightarrow \infty$, and we recover the perfectly flat, dispersionless Landau levels (2.22) from the homogeneous quantum Hall system, as shown in Fig. 2.3 (a). Note that the Landau levels are independent of the dispersive part $\hbar^2 k^2/(2m')$ at $k = 0$, and therefore we describe the chemical potential with respect to the sequence of Landau levels at $k = 0$. For example, the dashed lines in Fig. 2.3 (c) represent chemical potentials in between the first and second Landau level and in between the second and third Landau level, respectively.

In contrast to the flat Landau levels, we find an even number of dispersive modes, depending on the relative position of the chemical potential within the quantum Hall band structure [see Fig. 2.3 (c)]. Specifically, between the first and second Landau level there we find a single dispersive mode with positive group velocity

$$\mathbf{v}_k \equiv \frac{1}{\hbar} \frac{\partial \mathcal{E}_{n,k}}{\partial k} \mathbf{e}_x = \frac{\hbar k}{m'} \mathbf{e}_x \quad (2.39)$$

and another dispersive mode with negative group velocity $-\mathbf{v}_k$. These modes correspond to the chiral edge channels with the first one running along one edge in x direction, the other mode running along the opposite edge in $-x$ direction. By tuning the chemical potential so

that it lies in between the second and third Landau level, we find a total of four dispersing edge channels, two of which with positive group velocities and two with negative group velocities. Again, these modes are confined to the opposite edges of the quantum Hall system, and thus are of chiral nature. Apparently, for each Landau level below the chemical potential we obtain a set of two chiral edge channels propagating in opposite x and $-x$ directions. Since these edge channels always appear in pairs, there is no net current flowing through the quantum Hall sample in x direction [cf. Fig. 2.4 (b)].

So far we have motivated and explained the appearance of edge states in a quantum Hall sample, but we still have to relate their existence to the quantization of the Hall conductance, σ_{xy} . Let us consider a finite Hall voltage V_H across the quantum Hall sample in y direction, corresponding to the difference in the chemical potential between the two edges, $eV_H = \mu_R - \mu_L$ [see Fig. 2.4 (b)]. In the following, we assume that the charge transport occurs mainly at the edges and neglect the gradient in the chemical potential in the bulk. Moreover, we assume that the chiral edge modes are spatially separated by the width $L_y \gg l_B$ of sample, and are thus protected against backscattering from each other. Since the chemical potentials μ_L and μ_R of the external leads can be controlled independently, we can, in principle, control the number of chiral states at both edges independently. For example, by changing the chemical potentials appropriately, we can increase or decrease the number of edge channels and induce a net Hall current through the sample in x direction [see Fig. 2.4 (c)]. To calculate this Hall current we have to sum up the current densities of all occupied states in the quantum Hall sample:

$$I_x \equiv e \sum_{m=1}^{\nu} \int \frac{dk}{2\pi} (v_k^R n_F(\mathcal{E}_{m,k} - \mu_R) + v_k^L n_F(\mathcal{E}_{m,k} - \mu_L)), \quad (2.40)$$

where $v_k^{R,L}$ denotes the group velocity of right- and left-moving chiral edge states, and $n_F(\epsilon)$ is the Fermi distribution function. Note that we have assumed that there are ν Landau levels occupied in the bulk of the sample. In the zero temperature limit, only states below the chemical potential are occupied, i.e., $n_F(\epsilon - \mu) \rightarrow \Theta(\mu - \epsilon)$, and we can rewrite the integrand as a total derivative of the energy eigenvalues:

$$I_x = e \sum_{m=1}^{\nu} \int \frac{dk}{2\pi} \left(\frac{1}{\hbar} \frac{\partial \mathcal{E}_{m,k}}{\partial k} \Theta(\mu_R - \mathcal{E}_{m,k}) - \frac{1}{\hbar} \frac{\partial \mathcal{E}_{m,k}}{\partial k} \Theta(\mu_L - \mathcal{E}_{m,k}) \right) \quad (2.41a)$$

$$= \frac{e}{\hbar} \sum_{m=1}^{\nu} \int_{\mu_L}^{\mu_R} d\epsilon = \nu \frac{e}{\hbar} (\mu_R - \mu_L). \quad (2.41b)$$

Substituting the definition of the Hall voltage drop across the sample, $eV_H = \mu_R - \mu_L$, we obtain the current in x direction:

$$I_x = \nu \frac{e^2}{h} V_H. \quad (2.42)$$

As a consequence, we immediately see that the Hall conductance $\sigma_{xy} = I_x/V_H$ is quantized:

$$\sigma_{xy} = \nu \frac{e^2}{h} \quad \text{with} \quad \nu \in \mathbb{Z}. \quad (2.43)$$

Consequently, when the Fermi energy of the two-dimensional electron gas lies in a gap between two Landau levels, then the zero-temperature Hall conductance is an integer multiple of the conductance quantum $G_0 = e^2/h$. As mentioned before, this result does not depend on microscopic details of the sample such as the type of semiconducting material used or the quality

of the sample which in general enter the band structure \mathcal{E}_k . It is merely a result of the fact (i) that charges are quantized in multiples of the elementary charge e and (ii) that there is an integer number of conducting edge channels which cannot be changed adiabatically by tuning some external parameter. Hence, the quantum Hall effect is also robust against symmetry-breaking perturbations like disorder or crystal defects.

2.2.3 Chiral edge states in a tight-binding model

The appearance of current-carrying chiral edge states in a quantum Hall system actually does not depend on the specific boundary conditions of the sample. To show that explicitly, let us compute the spectrum of a tight-binding model on a square lattice and compare the results to the case of the harmonic confinement potential [cf. Eq. (2.31)]. In the following, we assume periodic boundary conditions in the x direction and open boundary conditions in the y direction. Note that the open boundary conditions can be considered as an effective, infinitely large hard-wall boundary potential, where edge channels are expected to appear when we switch on the magnetic field.

In the absence of an external magnetic field, the single-band tight-binding Hamiltonian for s orbital states on a two-dimensional square lattice takes the form

$$H = -t \sum_{\langle m,n \rangle} (c_m^\dagger c_n + \text{H.c.}). \quad (2.44)$$

Here, t is the **overlap parameter** or **hopping amplitude** for s orbitals, $\langle m, n \rangle$ denotes nearest-neighbor positions m and n , and c_n^\dagger and c_n are fermion creation and annihilation operators at site i satisfying the usual anti-commutation relations:

$$\{c_m, c_n\} = \{c_m^\dagger, c_n^\dagger\} = 0, \quad \{c_m, c_n^\dagger\} = \delta_{mn}, \quad (2.45)$$

with δ_{mn} the Kronecker delta. To account for the external magnetic field $\mathbf{B}(\mathbf{r}) = \text{curl } \mathbf{A}(\mathbf{r})$, we perform the so-called **Peierls substitution** which is the analogue of the minimal substitution $\mathbf{p} \rightarrow \mathbf{\Pi} = \mathbf{p} + e\mathbf{A}(\mathbf{r})$ for the electron gas on a discrete lattice:

$$t \rightarrow t' = t \exp\left(-\frac{ie}{\hbar} \int_{\mathbf{r}_i}^{\mathbf{r}_f} d\mathbf{r} \cdot \mathbf{A}(\mathbf{r})\right), \quad (2.46)$$

where the electron moves from the initial site \mathbf{r}_i to its final position \mathbf{r}_f . In the Landau gauge, $\mathbf{A}(\mathbf{r}) = -By\mathbf{e}_x$, the overlap parameters $t'_{x,y}$ describing hopping in the x and y directions acquire different phase factors:

$$t'_x = t \exp(iN_\Phi y), \quad t'_y = t. \quad (2.47)$$

Here, $N_\Phi \equiv \Phi/\Phi_0$ is the density of magnetic flux per unit cell, and $\Phi_0 = h/e$ denotes the quantum of magnetic flux. As a result, the tight-binding Hamiltonian in the presence of a uniform magnetic field is translationally invariant only in the x direction:

$$H = -t \sum_{\langle m,n \rangle} [\exp(iN_\Phi y) c_{mn}^\dagger c_{m-1,n} + c_{mn}^\dagger c_{m,n-1} + \text{H.c.}]. \quad (2.48)$$

Due to the translational invariance in x direction we may perform a Fourier decomposition of the creation and annihilation operators:

$$c_{mn}^\dagger = \frac{1}{\sqrt{N_x}} \sum_k e^{ikma} c_{kn}^\dagger, \quad c_{kn}^\dagger = \frac{1}{\sqrt{N_x}} \sum_m e^{-ikma} c_{kn}^\dagger, \quad (2.49)$$

where N_x is the number of unit cells in x directions and a the lattice constant. This leads to the following one-dimensional Bloch Hamiltonian:

$$\mathcal{H}(k) = -t \sum_n [\cos(ka + N_\Phi na) c_{kn}^\dagger c_{kn} + (c_{kn}^\dagger c_{k,n-1} + c_{kn}^\dagger c_{k,n+1})]. \quad (2.50)$$

Fig. 2.5 shows the energy eigenvalues or band structure of this tight-binding Hamiltonian as function of the momentum $\hbar k$ in x direction which is computed by exact diagonalization methods. Importantly, the hard-wall confinement by the sample edges due to the open boundary conditions in y direction leads to dispersive modes with positive and negative group velocities, ultimately resulting in the chiral edge states which are observed experimentally. Just as in the case of the harmonic confinement potential considered before, those edge states are very well localized at the boundaries of the sample [see Fig. 2.5 (c, d)]. We also find two chiral modes at the edges of the sample upon activating the second Landau level by increasing the chemical potential, etc. Note that the edge channel corresponding to the second Landau level is localized slightly more within the bulk than the first one.

2.3 The Hall conductance as a topological invariant

In this section, we discuss the role of topology in the quantum Hall effect. The basic idea can be easily understood by considering the single-particle Bloch functions of non-interacting electrons moving in a two-dimensional lattice. In the absence of a magnetic field, those Bloch functions are characterized by two quantum numbers, *i.e.*, the band index n and the two-dimensional crystal momentum \mathbf{k} . Since the wave function is periodic as a function of \mathbf{k} , the Brillouin zone is topologically equivalent to the two-dimensional torus T^2 . This idea of Bloch functions can be generalized to the so-called magnetic Bloch functions, where the momentum \mathbf{k} takes values in the magnetic Brillouin zone which is not the same as the crystalline Brillouin zone (see below). As we show below, it turns out that the Hall conductance σ_{xy} is an integer multiple of e^2/h , where the prefactor is the integral of some curvature over the torus T^2 . Mathematically, this quantity is a topological invariant which is known as the **first Chern class** of a so-called $U(1)$ bundle. This also implies that (i) the Hall conductance σ_{xy} is necessarily an integer multiple of e^2/h , and (ii) that this quantization is insensitive to experimental details such as the material used, the sample boundaries, disorder, electron-electron interactions, or an inhomogeneous magnetic field. In the following, we start with a review of the original argument due to Thouless, Kohmoto, Nightingale, and den Nijs¹⁴, before we proceed to rewrite the topological invariant in terms of a Berry phase. The following discussion is due to Watson³².

2.3.1 The argument by Thouless, Kohmoto, Nightingale, and den Nijs

In this section, we review the original version of the argument for a quantized Hall conductance which was given by Thouless *et al.* in 1982¹⁴. Let us start by considering a system of non-interacting electrons moving in the xy plane on a square lattice and subject to a uniform perpendicular magnetic field:

$$H = \frac{1}{2m} (\mathbf{p} + e\mathbf{A}(\mathbf{r}))^2 + U(\mathbf{r}) \quad \text{with} \quad U(x, y) = U(x + a, y) = U(x, y + a), \quad (2.51)$$

where $\mathbf{A}(\mathbf{r})$ is the vector potential, and the potential $U(\mathbf{r})$ simply represents the periodic potential of the square lattice. Then the Hall conductance can be calculated from a Kubo formula,

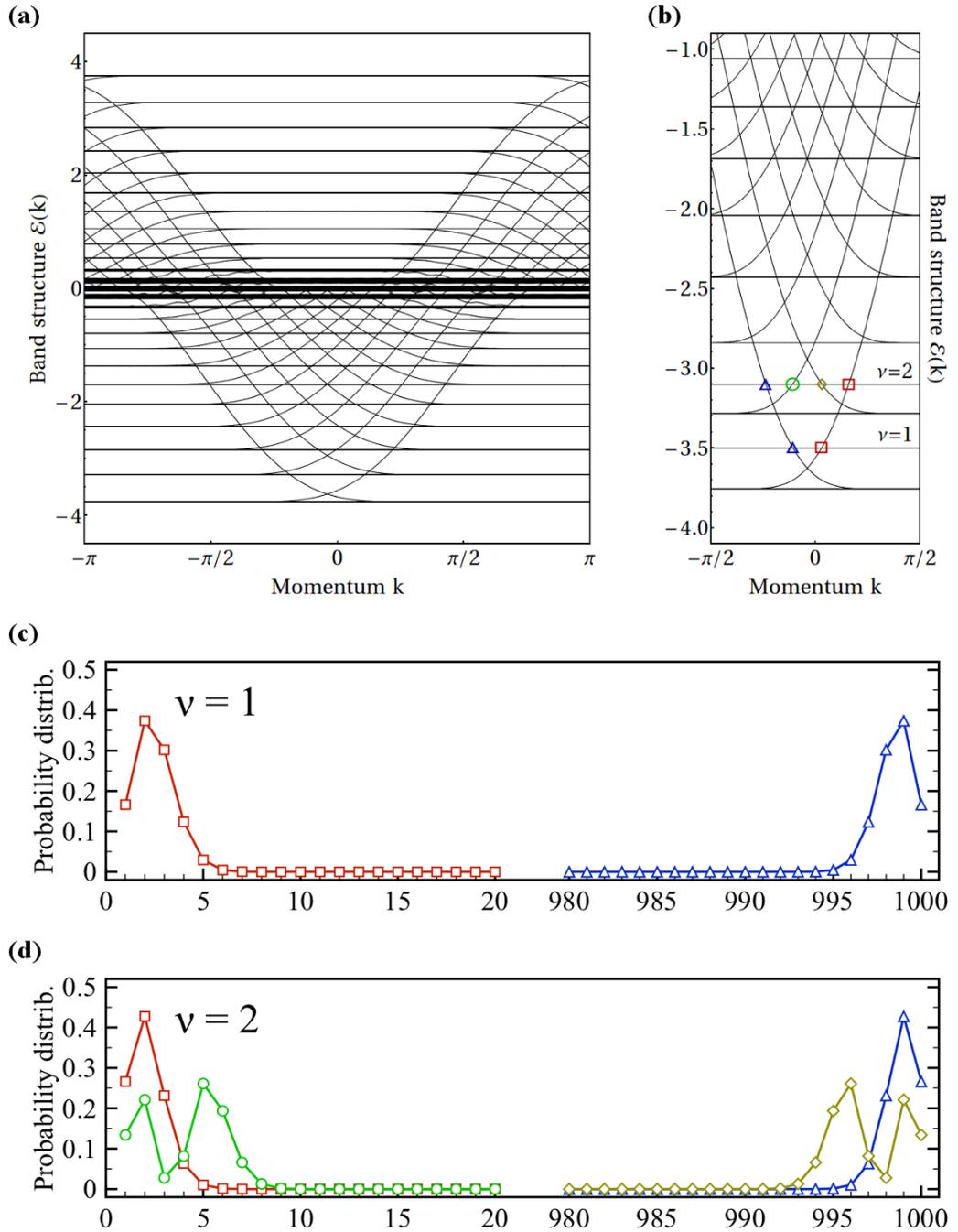


Figure 2.5. (a) Band structure of a quantum Hall system with infinite extension in x direction and finite width of $L_y/a = 1000$ sites in y direction. The flux density per unit cell is equal to $N_\Phi = 2\pi/25 \approx 0.251$, so that an electron has to hop around 25 plaquettes to enclose a single quantum of magnetic flux. In terms of length scales this is equivalent to the hierarchy $a \ll B \ll L_y$. The flat parts of the band structure correspond to the hierarchy of Landau levels, while the dispersing parts coincide with the chiral edge states at the boundaries. Upon increasing the chemical potential we observe an increasing number of left- and right-moving edge modes. (b) Zoom into the band structure close to the center of the Brillouin zone (the Γ point). Upon placing the chemical potential (gray lines) in between the Landau levels we observe an equal number of left- and right-moving chiral edge states. (c, d) Plot of the spatial distribution function for the $\nu = 1$ and $\nu = 2$ chiral edge modes with positive and negative group velocities. The edge states are well separated across the sample, leading to a vanishing overlap of the two wave functions within the bulk of the sample.

i.e., as response of the system to a small applied electric field in second-order perturbation theory³³:

$$\sigma_{xy} = i \frac{e^2}{\hbar \mathcal{A}} \sum_{\mathcal{E}_\alpha < \mathcal{E}_F < \mathcal{E}_\beta} \frac{v_x^{\alpha\beta} v_y^{\beta\alpha} - v_y^{\alpha\beta} v_x^{\beta\alpha}}{(\mathcal{E}_\alpha - \mathcal{E}_\beta)^2}. \quad (2.52)$$

Here, α, β label the single-particle eigenstates of the Hamiltonian with energies \mathcal{E}_α and \mathcal{E}_β , $v_{x,y}^{\alpha\beta}$ denote the matrix elements of the particle velocity operator in x and y directions, and \mathcal{A} is the 2D area of the system. Note that the energy denominator simply stems from a second-order process in which an electron below the Fermi surface ($\mathcal{E}_\alpha < \mathcal{E}_F$) is scattered to an empty state above the Fermi surface ($\mathcal{E}_\beta > \mathcal{E}_F$). Second, note that in the above expression for the Kubo formula we assume that the ground state is non-degenerate, and thus the denominator is always well-defined.

Let us consider the **magnetic unit cell** which is an enlarged unit cell in real space so that an integer multiple of the magnetic flux quantum Φ_0 passes through it [see Fig. 2.6 (a)]. For example, if the primitive lattice vectors of the magnetic unit cells take on the form $\mathbf{R} = n(q\mathbf{a}_1) + m\mathbf{a}_2$, then q magnetic flux quanta are in the unit cell formed by $q\mathbf{a}_1$ and \mathbf{a}_2 . Similar to the well-known Bloch states in a usual solid, the single-particle eigenstates $\psi_{n\mathbf{k}}(\mathbf{r})$ of the above Hamiltonian can be labelled by a band index n and a momentum \mathbf{k} in the so-called **magnetic Brillouin zone** which, for the example given above, takes on values restricted by $\mathbf{k} \in [0, 2\pi/(qa)] \times [0, 2\pi/a]$ [see Fig. 2.6 (b)]. Taking the translational invariance with respect to the magnetic unit cells into account, the corresponding **magnetic Bloch functions** $u_{n\mathbf{k}}(\mathbf{r})$ are normalized to unity,

$$\int d^2r |u_{n\mathbf{k}}(\mathbf{r})|^2 = 1 \quad \text{with} \quad u_{n\mathbf{k}}(\mathbf{r}) \equiv e^{-i\mathbf{k} \cdot \mathbf{r}} \psi_{n\mathbf{k}}(\mathbf{r}), \quad (2.53)$$

where the integral is taken over the magnetic unit cell. In terms of those Bloch functions, the matrix elements of the particle velocity operator can then be written as

$$\hbar \mathbf{v}^{\alpha\beta} = (\mathcal{E}_\beta - \mathcal{E}_\alpha) \langle u_\alpha | \nabla_{\mathbf{k}} u_\beta \rangle = (\mathcal{E}_\alpha - \mathcal{E}_\beta) \langle \nabla_{\mathbf{k}} u_\alpha | u_\beta \rangle. \quad (2.54)$$

As a consequence, the energy denominator in the Kubo formula (2.52) drops out upon substituting this expression for the components of the velocity operator $\mathbf{v}^{\alpha\beta}$. After rewriting the sum over states β as $\sum_{\mathcal{E}_\beta > \mathcal{E}_F} = \sum_{\mathcal{E}_\beta} - \sum_{\mathcal{E}_\beta < \mathcal{E}_F}$ it is then easy to show that the resulting expression for σ_{xy} can be written as

$$\sigma_{xy} = -i \frac{e^2}{\hbar \mathcal{A}} \sum_{\mathcal{E}_\alpha < \mathcal{E}_F} \left(\left\langle \frac{\partial u_\alpha}{\partial k_x} \middle| \frac{\partial u_\alpha}{\partial k_y} \right\rangle - \left\langle \frac{\partial u_\alpha}{\partial k_y} \middle| \frac{\partial u_\alpha}{\partial k_x} \right\rangle \right), \quad (2.55)$$

where in the last step we have used the completeness of the eigenstates to eliminate the sum over the eigenstates β . Note that the second term $\sum_{\mathcal{E}_\beta < \mathcal{E}_F}$ gives zero which can be seen by interchanging α and β eigenstates. The next assumption is that the Fermi energy lies in between a band gap, so that the sum extends over a set of filled bands α which can be rewritten as a momentum integral as follows:

$$\sum_{\mathcal{E}_\alpha < \mathcal{E}_F} \rightarrow \int d^2r \int \frac{dk_x}{2\pi/(qa)} \frac{dk_y}{2\pi/a}. \quad (2.56)$$

Due to the translational invariance of the magnetic Bloch functions, the matrix elements take values proportional to $\mathcal{A}/(qa^2)$ times the integral over the magnetic unit cell. Hence, the area

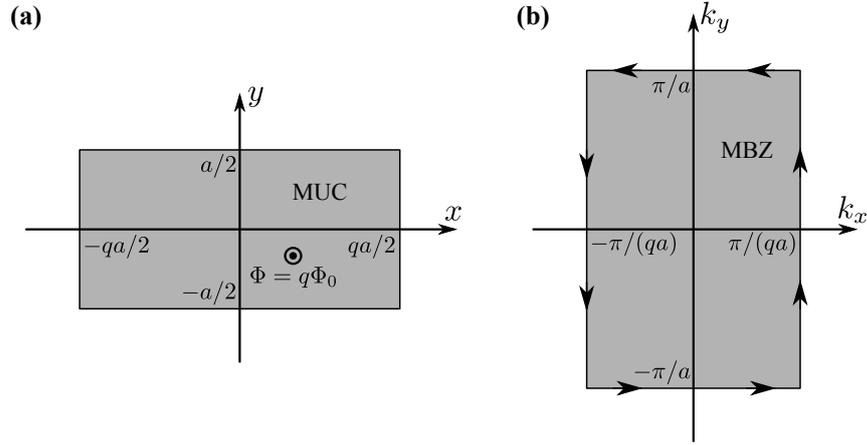


Figure 2.6. (a) The primitive lattice vectors of the magnetic unit cell (MUC), through which a total of q quanta of magnetic flux pass, is formed by the vectors qa_1 and a_2 , where $a_1 = ae_x$ and $a_2 = ae_y$ denote the primitive lattice vectors of the atomic square lattice. (b) Similar to the construction of the first Brillouin zone for solids, we can construct the magnetic Brillouin zone (MBZ) which, in this case, is characterized by the reciprocal lattice vectors $b_1 = (2\pi)/(qa)e_x$ and $b_2 = (2\pi/a)e_y$. The arrows indicate the direction of the line integral in Eq. (2.59). Figure after Ref. 32.

prefactor \mathcal{A} cancels, and we obtain the following momentum-space integral over magnetic Bloch functions as a result for the Hall conductance:

$$\sigma_{xy} = \frac{e^2}{h} \sum_{n \text{ occ.}} \int \frac{d^2k}{2\pi i} \left(\left\langle \frac{\partial u_{n\mathbf{k}}}{\partial k_x} \middle| \frac{\partial u_{n\mathbf{k}}}{\partial k_y} \right\rangle - \left\langle \frac{\partial u_{n\mathbf{k}}}{\partial k_y} \middle| \frac{\partial u_{n\mathbf{k}}}{\partial k_x} \right\rangle \right), \quad (2.57)$$

where the sum runs over occupied Landau levels, and the momentum integral is taken over the magnetic Brillouin zone only. Eq. (2.57) is the formula for the Hall conductance given by Thouless *et al.* in their original work¹⁴. At this point it is, however, not immediately clear that the above integral is quantized and takes only integer values.

As shown by Thouless *et al.*, the integrand appearing in Eq. (2.57) can be rewritten as the z component of the curl of some “fictitious” vector potential \mathcal{A} which will be defined later and whose interpretation is intimately related to the Berry phase of a spin 1/2. To show the integer quantization of the integral appearing in Eq. (2.57), we focus on the contribution of a single occupied band to the Hall conductance σ_{xy} . Starting from Eq. (2.57) we first insert a resolution of unity, $\mathbb{1} = \int d^2r |\mathbf{r}\rangle\langle\mathbf{r}|$, to rewrite the matrix elements in terms of complex-valued functions $u_{\mathbf{k}}(\mathbf{r})$ such that σ_{xy} takes the following form:

$$\sigma_{xy} = \frac{e^2}{h} \int d^2r \int \frac{d^2k}{2\pi i} \left(\frac{\partial u_{\mathbf{k}}(\mathbf{r})^*}{\partial k_x} \frac{\partial u_{\mathbf{k}}(\mathbf{r})}{\partial k_y} - \frac{\partial u_{\mathbf{k}}(\mathbf{r})^*}{\partial k_y} \frac{\partial u_{\mathbf{k}}(\mathbf{r})}{\partial k_x} \right). \quad (2.58)$$

We can write the integrand as the z component of the curl of some vector field, and then use Stokes’ theorem to rewrite the integral as a line integral around the magnetic Brillouin zone [see Fig. 2.6 (b)]:

$$\sigma_{xy} = \frac{e^2}{h} \int d^2r \int \frac{d^2k}{2\pi i} [\nabla_{\mathbf{k}} \times (u_{\mathbf{k}}(\mathbf{r})^* \nabla_{\mathbf{k}} u_{\mathbf{k}}(\mathbf{r}))]_z \quad (2.59a)$$

$$= \frac{e^2}{h} \int d^2r \oint \frac{d\mathbf{k}}{2\pi i} \cdot (u_{\mathbf{k}}(\mathbf{r})^* \nabla_{\mathbf{k}} u_{\mathbf{k}}(\mathbf{r})). \quad (2.59b)$$

The remaining line integral can be computed in a straightforward way. First, note that the wave functions $u_{\mathbf{k}}(\mathbf{r})$ can differ only by a phase factor $\gamma(\mathbf{k})$ between corresponding points in the magnetic Brillouin zone shown in Fig. 2.6 (b), *i.e.*, on the horizontal (vertical) lines we have

$$u_{k_x, \pi/a}(\mathbf{r}) = e^{i\gamma(\mathbf{k})} u_{k_x, -\pi/a}(\mathbf{r}), \quad (2.60a)$$

$$u_{\pi/(qa), k_y}(\mathbf{r}) = e^{i\gamma(\mathbf{k})} u_{-\pi/(qa), k_y}(\mathbf{r}). \quad (2.60b)$$

Note that $\gamma(\mathbf{k})$ is independent of the position \mathbf{r} . The line integral along the boundary of the magnetic Brillouin zone then gives the following contribution to the Hall conductance:

$$\sigma_{xy} = \frac{e^2}{h} \oint \frac{d\mathbf{k}}{2\pi i} \cdot (i\nabla_{\mathbf{k}}\gamma(\mathbf{k})) = \frac{e^2}{h} \oint \frac{d\mathbf{k}}{2\pi} \cdot \nabla_{\mathbf{k}}\gamma(\mathbf{k}), \quad (2.61)$$

where we have used the fact that the magnetic Bloch functions are taken to be normalized to unity [cf. Eq. (2.53)]. Therefore, the total line integral is proportional to the total change in the phase $\gamma(\mathbf{k})$ of the wave function around the boundary of the magnetic Brillouin zone, and since that phase must be an integer multiple of 2π , this finally shows that the Hall conductance must be quantized:

$$\sigma_{xy} = \frac{e^2}{h} \oint \frac{d\mathbf{k}}{2\pi} \cdot \nabla_{\mathbf{k}}\gamma(\mathbf{k}) = \frac{e^2}{h} \times \nu. \quad (2.62)$$

Most importantly, ν is an integer number. That result can now be easily generalized to a set of n occupied bands:

$$\sigma_{xy} = \frac{e^2}{h} \sum_{n \text{ occ.}} \oint \frac{d\mathbf{k}}{2\pi} \cdot \nabla_{\mathbf{k}}\gamma_n(\mathbf{k}) = \nu \frac{e^2}{h} \quad (2.63)$$

with the integer ν known as the **TKNN invariant**. This concludes our review of the proof given by Thouless *et al.* in their original work for the quantized value of the Hall conductance¹⁴.

To sum up, we started from a Kubo formula (2.52) for the Hall conductance of a non-interacting two-dimensional electron gas which is only valid (i) when the ground state is non-degenerate and (ii) when linear response theory is applicable, *i.e.*, when the electric field is smaller than the energy gap and can be treated as a small perturbation. This expression was then expressed in terms of magnetic Bloch functions $u_{n\mathbf{k}}(\mathbf{r})$, but very little information was actually needed to rewrite the Kubo formula in the form (2.57) given in the original work by Thouless *et al.* (Ref. 14). The only important fact used was that those Bloch functions can be labelled by a band index n and a momentum vector \mathbf{k} in the magnetic Brillouin zone. Moreover, we made use of the fact that the Bloch functions have a well-defined phase relationship when taking the line integral along the boundary of the magnetic Brillouin zone. This ultimately lead us the conclusion that the Hall conductance σ_{xy} of a non-interacting electron gas is necessarily an integer multiple of the conductance quantum e^2/h .

The weakness of the argument is, however, that the above reasoning only applies in clean systems, and one might ask whether disorder could destroy the translational symmetry with respect to the magnetic unit cells, so that the use of magnetic Bloch functions $u_{n\mathbf{k}}(\mathbf{r})$ is no longer applicable. The generalization of the argument by Thouless *et al.* was given by Niu, Thouless, and Wu in another seminal work in 1985³⁴. This argument only assumes that the many-body ground state of the system is non-degenerate and that an energy gap exists at the Fermi level, so that all bands below the Fermi energy are completely filled. Niu *et al.* then showed that the Hall conductance σ_{xy} can be expressed as an integral similar to that one given in Eq. (2.57), and thus σ_{xy} is quantized in integer multiples of e^2/h .

2.3.2 The TKNN invariant as a topological invariant

Mathematically, the integral appearing in Eq. (2.57) is indeed a **topological invariant**. It was first recognized by Avron, Seiler, and Simon in 1983 that this integral is the **first Chern class** of a so-called $U(1)$ fiber bundle on the base manifold of a torus T^2 (Refs. 35,36). In the case of non-interacting electrons, the fibers are the magnetic Bloch functions $|u_{n\mathbf{k}}\rangle$, and the torus T^2 corresponds to the magnetic Brillouin zone. On the other hand, for interacting electrons the integer quantization of the Hall conductance was previously shown by Laughlin in a seminal work, where he viewed the quantum Hall effect as a quantum charge pump, exploiting the fact that the Hamiltonian must be the same when an integer multiple of flux quanta has passed through the system²². Nevertheless, the recognition of the Hall conductance as a topological invariant is not just a mathematical formality, but also of physical relevance as it explains the robustness of the quantization of the Hall conductance against many kinds of perturbations.

To gain a better understanding of the relation between the Hall conductance and the first Chern class of non-interacting electrons, let us consider the Hall conductance of a single band in units of e^2/h [cf. Eq. (2.59)]:

$$\frac{\sigma_{xy}}{e^2/h} = \int \frac{d^2k}{2\pi i} \nabla_{\mathbf{k}} \times \left(\int d^2r u_{\mathbf{k}}(\mathbf{r})^* \nabla_{\mathbf{k}} u_{\mathbf{k}}(\mathbf{r}) \right)_z. \quad (2.64)$$

Now let us define a fictitious vector potential $\mathcal{A}(\mathbf{k})$ in momentum space by

$$\mathcal{A}(\mathbf{k}) \equiv i \int d^2r u_{\mathbf{k}}(\mathbf{r})^* \nabla_{\mathbf{k}} u_{\mathbf{k}}(\mathbf{r}) = i \langle u_{\mathbf{k}} | \nabla_{\mathbf{k}} u_{\mathbf{k}} \rangle. \quad (2.65)$$

Here, two aspects are noteworthy: First, the vector potential $\mathcal{A}(\mathbf{k})$ is formally defined in terms of the magnetic Bloch functions $|u_{\mathbf{k}}\rangle$ in momentum space, and thus it has no connection at all to any electric or magnetic field in the sample. Second, $\mathcal{A}(\mathbf{k})$ is also known as the **Berry vector potential**, and Eq. (2.64) shows that the Hall conductance σ_{xy} can be expressed as the curl of $\mathcal{A}(\mathbf{k})$:

$$-\frac{\sigma_{xy}}{e^2/h} = \int \frac{d^2k}{2\pi} (\nabla_{\mathbf{k}} \times \mathcal{A}(\mathbf{k}))_z = \int \frac{d^2k}{2\pi} \mathcal{F}_z(\mathbf{k}) = \frac{\phi}{2\pi}, \quad (2.66)$$

where $\mathcal{F}(\mathbf{k}) \equiv \nabla_{\mathbf{k}} \times \mathcal{A}(\mathbf{k})$ is the so-called **Berry curvature**, and $\phi \equiv \oint d\mathbf{k} \cdot \mathcal{A}(\mathbf{k})$ is the well-known **Berry phase**. Here, we have used that the integral of $\nabla_{\mathbf{k}} \times \mathcal{A}(\mathbf{k})$ over the magnetic Brillouin zone can be converted into a line integral of the Berry vector potential along the boundary of the magnetic Brillouin zone. Note, however, that this Brillouin zone is topologically equivalent to a torus T^2 , and hence the Hall conductance necessarily vanishes if one can smoothly define the vector potential $\mathcal{A}(\mathbf{k})$ over the entire torus. Conversely, this implies that a necessary condition for a non-trivial Hall conductance is that it is impossible to find magnetic Bloch functions $u_{\mathbf{k}}(\mathbf{r})$, so that $\mathcal{A}(\mathbf{k})$ is smoothly defined on the torus. Therefore, one cannot apply Stokes' theorem globally over the whole magnetic Brillouin zone.

Mathematically, this observation can be reformulated more strictly using topology. Since the magnetic Bloch functions are only defined up to a phase factor they form a $U(1)$ fiber bundle over the base manifold of the magnetic Brillouin zone which is topologically equivalent to the torus T^2 . As a consequence, the fictitious vector potential $\mathcal{A}(\mathbf{k})$ possesses a $U(1)$ gauge freedom. The integral over the Berry vector potential $\mathcal{A}(\mathbf{k})$ is the **Chern number** which is a topological invariant of the $U(1)$ fiber bundle. Although $\mathcal{A}(\mathbf{k})$ has a gauge degree of freedom, the Chern number is invariant under gauge transformations, and thus the Hall conductance is well-defined and an integer multiple of e^2/h .

This concludes the discussion of the topological aspects of the quantum Hall effect. At this point, the connection between the quantized value of the Hall conductance and the first Chern number (or TKNN invariant), which is a measure for the non-trivial structure of a $U(1)$ fiber bundle, should have become clear. In the following, we will often use a similar terminology when it comes to the discussion of the topological properties of some model Hamiltonian.

2.4 Quantum Hall effect without Landau levels: Haldane model

The quantum Hall effect of a two-dimensional electron gas is usually associated with the presence of an applied external magnetic field which gives rise to the Landau level spectrum (2.22). However, it was shown in 1988 by Haldane that one can observe a quantum Hall effect even in a system without an external magnetic field, but with broken time-reversal symmetry². In general, the scenario considered in the Haldane model involves a two-dimensional semimetal with a degeneracy of the valence and conduction bands at isolated points in the Brillouin zone. If inversion symmetry is broken, a gap opens, and the system becomes a normal insulator with $\sigma_{xy} = 0$. However, if a gap opens due to the breaking of time-reversal symmetry, the system becomes an integer quantum Hall insulator with quantized Hall conductance σ_{xy} . Nevertheless, if both perturbations are present, the two gaps compete, so that the relative strength of the two symmetry-breaking terms determines whether the system exhibits a quantum Hall state or not.

Although Haldane's model was originally defined on the honeycomb lattice², we consider, for simplicity, the following two-band model of s orbital states on the two-dimensional square lattice:

$$H = -t \sum_{i,j} \left(\Psi_{ij}^\dagger \frac{\sigma_z + i\sigma_x}{2} \Psi_{i+1,j} + \Psi_{ij}^\dagger \frac{\sigma_z + i\sigma_y}{2} \Psi_{i,j+1} + \text{H.c.} \right) + m \sum_{i,j} \Psi_{ij}^\dagger \sigma_z \Psi_{ij}. \quad (2.67)$$

Here, t is the **overlap parameter** or **hopping amplitude**, m is the **tuning parameter** or fermion mass, and Ψ is a two-component spinor. Moreover, σ_x , σ_y , and σ_z are the usual Pauli matrices in standard notation:

$$\sigma_x = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad \sigma_y = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \quad \sigma_z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \quad (2.68)$$

To diagonalize this Hamiltonian we may perform a Fourier decomposition of the two-component spinors Ψ and rewrite the model in momentum space, leading to the following Bloch Hamiltonian:

$$\mathcal{H}(\mathbf{k}) = \mathbf{d}(\mathbf{k}) \cdot \boldsymbol{\sigma}, \quad (2.69)$$

where, for brevity, we have introduced the vector $\mathbf{d}(\mathbf{k})$ by

$$\mathbf{d}(\mathbf{k}) \equiv (-t \sin(k_x a), -t \sin(k_y a), m - t \cos(k_x a) - t \cos(k_y a))^T. \quad (2.70)$$

The band structure of the Haldane model is obtained by straightforward diagonalization of the Bloch Hamiltonian, leading to the following conduction and valence bands:

$$\mathcal{E}_\pm(\mathbf{k}) = \pm |\mathbf{d}(\mathbf{k})| = \pm \sqrt{\sum_{j=1}^3 d_j(\mathbf{k})^2}. \quad (2.71)$$

Fig. 2.7 shows the bulk band structure for $m/t = \pm 0.1$ for a system with periodic boundary conditions in x and y directions as function of the crystal momentum \mathbf{k} .

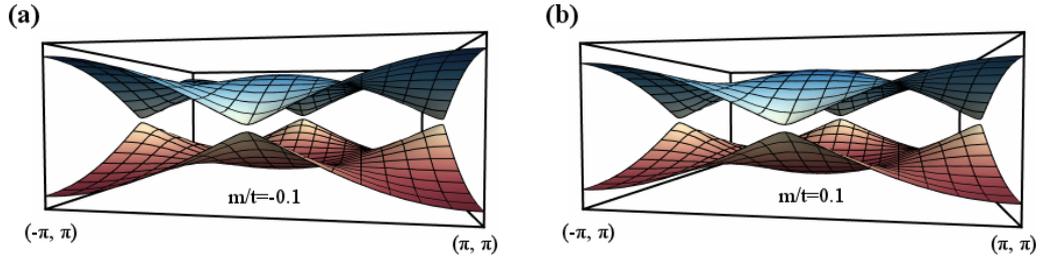


Figure 2.7. Bulk band structure $\mathcal{E}_{\pm}(\mathbf{k})$ of the Haldane model for $m/t = 0.1$ [panel (a)] and $m/t = -0.1$ [panel (b)]. Obviously, the band structure is fully gapped with the size of the bulk band gap given by $\mathcal{E}_{\text{gap}}/t = 2|m/t| = 0.2$ which allows for the calculation of the Chern number. A close inspection reveals only minor differences between the two band structures for $m/t = \pm 0.1$. Nevertheless, the two band structures are characterized by different Chern numbers $n = -1$ and $n = 1$ for $m/t = \pm 0.1$ in the left (right) panel, respectively.

2.4.1 The first Chern number

Despite the seemingly simple Hamiltonian (2.69), this model exhibits an unusual phase diagram consisting of three phases that may be distinguished by some non-local order parameter (see Fig. 2.9). It turns out that this non-local order parameter is identical to the **first Chern number** ν , similar to the order parameter describing the quantized conductance σ_{xy} in the quantum Hall effect. In terms of the vector $\mathbf{d}(\mathbf{k})$, this topological invariant ν can be written as

$$\nu = \frac{1}{4\pi} \int dk_x dk_y \hat{\mathbf{d}}(\mathbf{k}) \cdot (\partial_{k_x} \hat{\mathbf{d}}(\mathbf{k}) \times \partial_{k_y} \hat{\mathbf{d}}(\mathbf{k})) \in \mathbb{Z}, \quad (2.72)$$

where $\hat{\mathbf{d}}(\mathbf{k}) \equiv \mathbf{d}(\mathbf{k})/|\mathbf{d}(\mathbf{k})|$ is the unit vector in the direction of $\mathbf{d}(\mathbf{k})$. Geometrically, the Chern number is equal to the number of times the unit vector $\hat{\mathbf{d}}(\mathbf{k})$ wraps around the unit sphere as we go through the whole Brillouin zone. Furthermore, note that the above construction of a winding number has a deep relation to the second homotopy group of the unit sphere in three dimensions, $\nu \in \pi_2(S^2) = \mathbb{Z}$. For example, the two band structures shown in Fig. 2.7 look quite similar on either side of the quantum critical point $m/t = 0$, but the Chern number takes on different values $\nu = 1$ or $\nu = -1$ for $-2 < m/t < 0$ and $0 < m/t < 2$, respectively. Although the electronic band structures plotted in Fig. 2.7 do not immediately reveal their topological properties, we can understand the different Chern numbers by analyzing directly the vector $\hat{\mathbf{d}}(\mathbf{k})$ which enters Eq. (2.72).

Fig. 2.8 shows an illustration of the vector $\hat{\mathbf{d}}(\mathbf{k})$ as function of the crystal momentum \mathbf{k} for different values of the tuning parameter m . The color code in that figure is chosen such that arrows pointing up (down) are indicated by red (blue) color, while arrows lying in the equatorial plane are shown in green color. The first Brillouin zone $[-\pi/a, \pi/a] \times [-\pi/a, \pi/a]$ is indicated by the black square as a guide to the eye. Obviously, in the topologically trivial phases for $|m/t| > 2$ the normalized vector $\hat{\mathbf{d}}(\mathbf{k})$ points either up or down, but does not show any winding [cf. Fig. 2.8 (a, d)]. In the topologically non-trivial phases, however, we observe a skyrmion configuration at $\mathbf{k} = (\pi/a, \pi/a)^T$ with Chern number $\nu = 1$ for $m/t = -1$ and an anti-skyrmion configuration at $\mathbf{k} = (0, 0)^T$ with opposite Chern number $\nu = -1$ for $m/t = 1$ [cf. Fig. 2.8 (b, c)]. Note that in order to change the Chern number ν , the system has to undergo a quantum phase transition, where the bulk band gap vanishes, and the unit vector $\hat{\mathbf{d}}(\mathbf{k})$ becomes singular. In the Haldane model, this can happen at $m/t = \pm 2$ and $m/t = 0$ (see Fig. 2.9) at one of the high-symmetry points in the Brillouin zone, i.e., $\Gamma = (0, 0)^T$, $M = (\pi/a, 0)^T$ or $M = (0, \pi/a)^T$, and $R = (\pi/a, \pi/a)^T$.

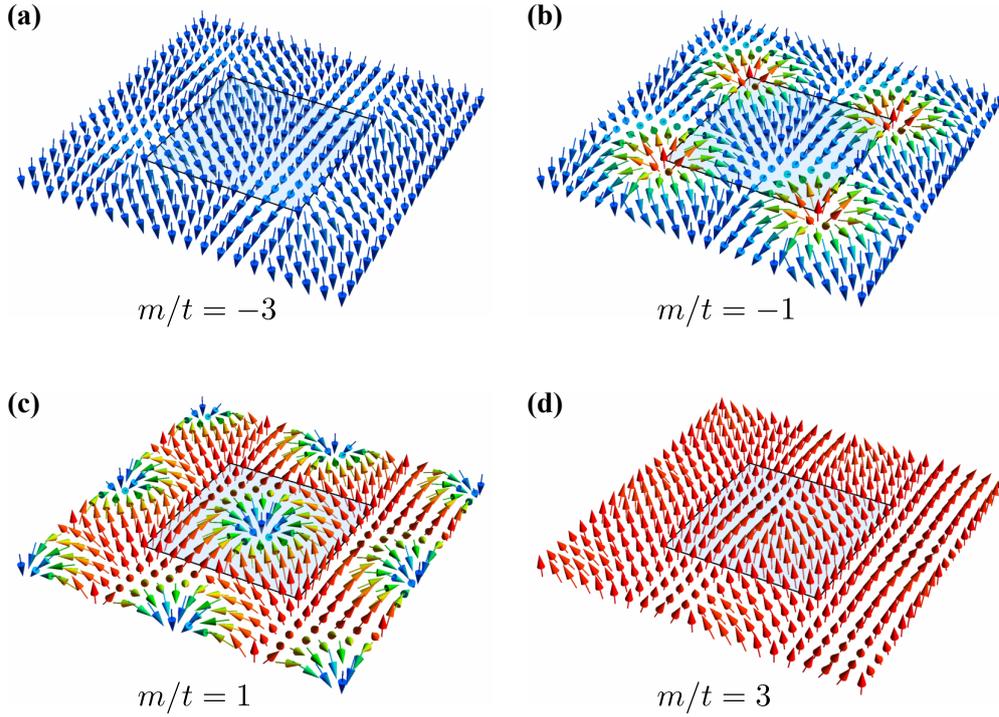


Figure 2.8. Illustration of the winding of the unit vector $\hat{\mathbf{d}}(\mathbf{k})$ as function of the two-dimensional crystal momentum \mathbf{k} , where $\mathbf{d}(\mathbf{k})$ is defined by Eq. (2.70) in the main text. The color code is chosen such that red (blue) arrows point up (down), while green arrows lie in the equatorial plane, and the black square indicates the first Brillouin zone $[-\pi/a, \pi/a] \times [-\pi/a, \pi/a]$ of the underlying square lattice. Panels (a) and (d) show the absence of a winding of $\hat{\mathbf{d}}(\mathbf{k})$ in the usual band insulator phase for $|m/t| > 2$. In contrast, for $|m/t| < 2$ we observe a non-trivial winding of $\hat{\mathbf{d}}(\mathbf{k})$ which changes at the quantum critical point $m/t = 0$, as shown in panels (b) and (c). Moreover, for $m/t = -1$ we observe a skyrmion configuration of $\hat{\mathbf{d}}(\mathbf{k})$ with winding number $\nu = 1$ located at the corner $\mathbf{k} = (\pi/a, \pi/a)^T$ of the Brillouin zone, while for $m/t = 1$ there exists an anti-skyrmion located at the center $\mathbf{k} = (0, 0)^T$ with winding number $\nu = -1$.

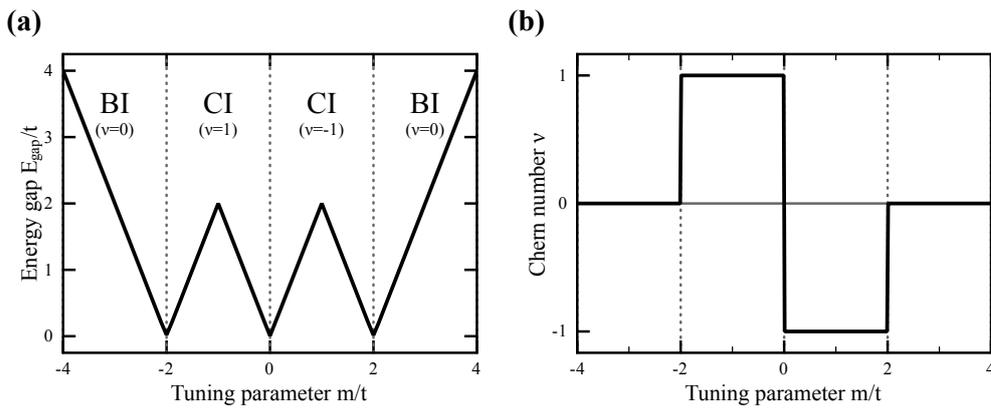


Figure 2.9. (a) Phase diagram of the Haldane model (2.67) as function of the tuning parameter m . For $|m/t| > 2$, the system describes an ordinary band insulator (BI) with Chern number $\nu = 0$, while for $|m/t| < 2$ we have a Chern insulator (CI) with Chern number $\nu = \pm 1$ for negative (positive) m . (b) Plot of the Chern number ν as function of m , calculated numerically from Eq. (2.72) for the Chern insulator (2.67). To change the topological sector, the system has to undergo a quantum phase transitions at $m/t = \pm 2$ or $m = 0$, where the bulk band gap vanishes.

2.4.2 The appearance of edge states in the Haldane model

To complete the discussion of the Haldane model, we discuss the effect of boundaries on the band structure. Let us consider the tight-binding Hamiltonian (2.67) as real-space version of the Haldane model with periodic boundary conditions in x direction and open boundary conditions in y direction. Due to the translational invariance, we may perform a Fourier decomposition of the spinors Ψ with respect to the x direction, leading to an effective one-dimensional Bloch Hamiltonian $\mathcal{H}(k)$:

$$\mathcal{H}(k) = \sum_j \psi_j^\dagger [-t \sin(ka) \sigma_x + (m - t \cos(ka)) \sigma_z] \psi_j - t \sum_j \left(\psi_j^\dagger \frac{\sigma_z + i\sigma_y}{2} \psi_{j+1} + \text{H.c.} \right), \quad (2.73)$$

where k denotes the conserved momentum in x direction. The band structure for this one-dimensional Hamiltonian can be calculated numerically by exact diagonalization techniques. As a result, we find that the spectrum is a projection of the bulk band structure (2.71) onto the conserved momentum k within the bulk Brillouin zone, shown in Fig. 2.10 for $m/t = \pm 0.1$. However, a close inspection of the band structure reveals additional modes traversing the bulk band gap which basically invalidate the picture of an insulator away from the quantum critical points $m/t = \pm 2$ and $m/t = 0$. Furthermore, those modes cease to exist when going into the ordinary insulating phase ($|m/t| > 2$) with Chern number $\nu = 0$. Similar to the chiral edge states in the quantum Hall state, we may identify those modes as topologically protected edge states being located at the boundaries, where the Chern number changes from a non-trivial value $\nu = \pm 1$ to the trivial case $\nu = 0$, and those edge states cross in a Dirac-like fashion at the time-reversal invariant momenta $k = 0$ or $k = \pi$ (see Fig. 2.10)

2.5 Quantum spin Hall insulator

The quantum Hall state provided the first example of a quantum state that cannot be characterized by a local order parameter theory, but shows an **intrinsic topological order** which is stable even without any symmetry. Recently, a new class of topological states has emerged which are called **quantum spin Hall insulators** or **topological insulators**³⁷. The quantum spin Hall insulator has been theoretically predicted for graphene^{3,4}, and has first been experimentally observed in mercury telluride quantum wells^{5,6,8}. This discovery has motivated the study of the so-called **symmetry-protected topological order**. In contrast to systems with intrinsic topological order, the symmetry-protected topological states are only distinguishable from trivial disordered phases when a certain symmetry is preserved, most notably time-reversal invariance.

While the quantum spin Hall state and the quantum Hall state are quite similar and possess a lot of common properties, they also differ in important ways, most importantly in the presence or absence of time-reversal symmetry, respectively. Since the Hall conductivity σ_{xy} is odd under time-reversal, the chiral edge states in the quantum Hall state can only be observed when time-reversal symmetry is broken, for example by a magnetic field. As discussed before, time-reversal symmetry is also broken in the Haldane model (2.67) by an intrinsic staggered magnetic field with no net flux through the unit cell, and this model also exhibits a quantized Hall conductance. Starting from the Haldane model one can construct a toy Hamiltonian for a quantum spin Hall insulator by combining two time-reversed copies of that model in such a way that no net magnetic field (neither intrinsic nor externally applied) penetrates the sample. As a result, this Hamiltonian preserves time-reversal symmetry, because time-reversal flips both the spin-up and spin-down states as well as the sign of σ_{xy} . Moreover, in an applied electric

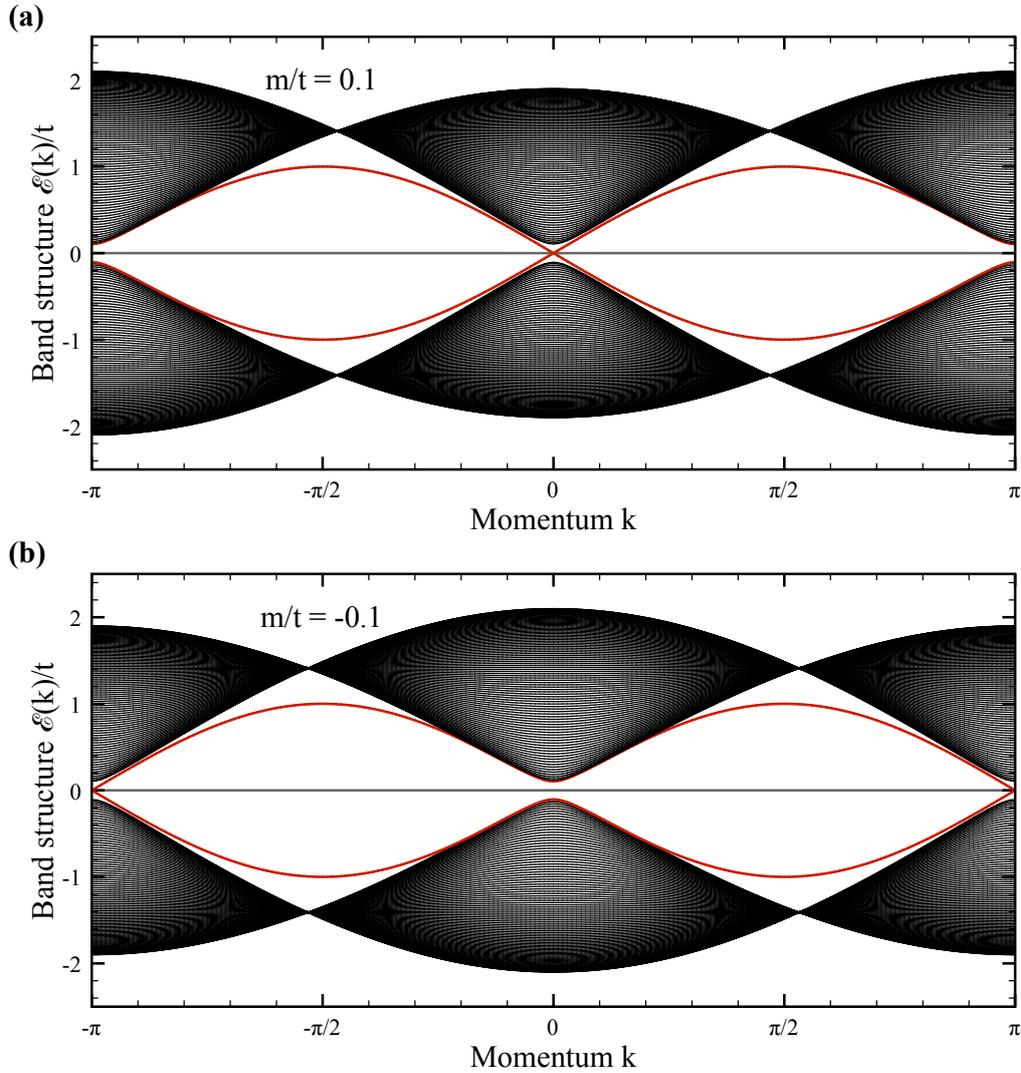


Figure 2.10. Band structure for the semi-infinite Haldane model (2.73) with periodic boundary conditions in x direction and open boundary conditions in y direction. While most bands are related to the bulk band structure by straightforward projection of the bulk bands along the y direction, there exist boundary modes (indicated by red color) traversing the band gap. The location of the crossing point of the helical edge states changes from $k = 0$ for $m/t > 0$ [panel (a)] to $k = \pi/a$ with $a = 1$ for $m/t < 0$ [panel (b)].

field, the spin-up and spin-down electrons give rise to spin-polarized Hall currents flowing in opposite directions which are referred to as **helical edge states**, in analogy to the relation between the spin and the momentum of a particle known as **helicity** from relativistic quantum mechanics. Although the TKNN invariant is zero, there is, however, another topological \mathbb{Z}_2 **invariant** ν for the quantum spin Hall insulator which takes two possible values, $\nu \in \{0, 1\}$, and which distinguishes between two topologically distinct classes. The fact that there are only two classes can be understood by considering the edge states in a quantum spin Hall insulator. Being comprised of two copies of the Haldane model which are time-reversal partners, there are always pairs of counterpropagating edge states with opposite spin present at each edge or surface which we refer to as moving in “forward” and “backward” directions, respectively. For a single pair of those edge states, backscattering by a non-magnetic impurity is forbidden due to time-reversal symmetry. Kane and Mele showed that the reflection amplitude from a non-magnetic impurity is odd under time-reversal, because it basically involves flipping the electron’s spin³. Unless time-reversal symmetry is broken by, for example, a magnetic impurity, an incident electron has to be perfectly transmitted across the impurity region. In that sense, the helical edge states are **topologically protected** by time-reversal symmetry.

2.5.1 Construction of the \mathbb{Z}_2 invariant for 2D topological insulators

Over the last couple of years, a large number of equivalent mathematical formulations of the \mathbb{Z}_2 invariant ν have been developed^{3,8,11,38–46}. The original approach by Fu and Kane³⁸ is formulated in terms of the occupied Bloch functions $|u_m(\mathbf{k})\rangle$. From quantum mechanics we recall that time-reversal symmetry is represented by an anti-unitary operator $\hat{\Theta} = \exp(\frac{i\pi}{\hbar} S_y) \mathcal{K}$, where S_y is the spin operator, and \mathcal{K} denotes complex conjugation. Moreover, for spin 1/2 electrons we know that $\hat{\Theta}^2 = -\mathbb{1}$, where $\mathbb{1}$ is the identity operator. Note that this leads to the important constraint that all eigenstates of a time-reversal invariant Hamiltonian are at least two-fold degenerate (**Kramers’ theorem**). Starting from the Bloch states $|u_m(\mathbf{k})\rangle$ one can analyze their properties under time-reversal by defining a unitary matrix w as:

$$w_{mn}(\mathbf{k}) \equiv \langle u_m(\mathbf{k}) | \hat{\Theta} | u_n(-\mathbf{k}) \rangle. \quad (2.74)$$

Due to time-reversal symmetry, the matrix elements at \mathbf{k} and $-\mathbf{k}$ are related by

$$w(\mathbf{k})^T = -w(-\mathbf{k}). \quad (2.75)$$

In two dimensions, the Brillouin zone contains four **time-reversal invariant momenta** (TRIMs) which are denoted by Γ_a with $a = 1, \dots, 4$:

$$\Gamma_1 = (0, 0)^T, \quad \Gamma_2 = (\pi, 0)^T, \quad \Gamma_3 = (0, \pi)^T, \quad \Gamma_4 = (\pi, \pi)^T. \quad (2.76)$$

At these four points $w(\Gamma_a)$ is skew-symmetric since

$$w(\Gamma_a)^T = -w(-\Gamma_a) = -w(\Gamma_a). \quad (2.77)$$

Mathematically, the determinant of a skew-symmetric matrix can be written as the square of the Pfaffian, a polynomial in the matrix entries. This allows for the definition of time-reversal parities δ_a by

$$\delta_a \equiv \frac{\text{Pf } w(\Gamma_a)}{\sqrt{\det w(\Gamma_a)}} = \pm 1. \quad (2.78)$$

Provided the gauge of the Bloch states $|u_m(\mathbf{k})\rangle$ is chosen continuously throughout the whole Brillouin zone, one can define the branch of the square root uniquely. Given a surface normal

vector \mathbf{G} , we can identify pairs of time-reversal invariant momenta by $\Gamma_{a1} - \Gamma_{a2} = \mathbf{G}/2$, and the so-called “time-reversal polarizations” π_a (see Ref. 38) associated with that surface can be written as

$$\pi_a \equiv \delta_{a1}\delta_{a2} = \pm 1. \quad (2.79)$$

For a topologically non-trivial surface state an odd number of states is required which have an odd parity under time-reversal. This allows to define the \mathbb{Z}_2 invariant ν by the following relation:

$$(-1)^\nu \equiv \prod_{a=1}^4 \delta_a. \quad (2.80)$$

In the two-dimensional quantum spin Hall phase with $\nu = 1$ the product of two time-reversal parity eigenvalues is equal to $\pi_a\pi_b = -1$, while all other combinations are equal to $+1$.

If the Hamiltonian contains additional symmetries, the calculation of the \mathbb{Z}_2 invariant ν may be simplified. For example, if the Hamiltonian has inversion symmetry, the Bloch states $|u_m(\mathbf{k})\rangle$ are also parity eigenstates at the special points Λ_a with eigenvalues $\chi_m(\Lambda_a) = \pm 1$. In that case, the polarizations δ_a entering the definition of ν [Eq. (2.80)] have been shown to be equal to the products of parity eigenvalues of the occupied bands⁷:

$$\delta_a = \prod_{m \text{ occ.}} \chi_m(\Lambda_a). \quad (2.81)$$

If, however, the Hamiltonian conserves the total spin S_z , then one can define two independent Chern numbers n_\uparrow and n_\downarrow for the spin-up and spin-down electrons which describes the quantized spin Hall conductance of the system^{3,4,39,47}, where time-reversal invariance implies that $n_\uparrow + n_\downarrow = 0$. Nevertheless, the \mathbb{Z}_2 invariant can be identified with the parity of the spin Chern number:

$$\nu = \frac{n_\uparrow - n_\downarrow}{2} \pmod{2}. \quad (2.82)$$

However, in the presence of terms that do not conserve S_z , an experimental reality, the spin Chern number loses its meaning, but the \mathbb{Z}_2 invariant retains its value and characterized the quantum spin Hall phase.

The above construction of the \mathbb{Z}_2 invariant ν basically relies on the fact that the model is time-reversal invariant. Kane and Mele showed that as a consequence of that symmetry the edge states in the quantum spin Hall insulator phase are robust against the effects of weak interactions, even when spin conservation is broken^{3,4}. Nevertheless, for strong interactions an electronic instability opening up a band gap is expected, and the resulting state breaks time-reversal symmetry^{38,48,49}.

2.5.2 Generalization to 3D topological insulators

The above construction of a topological \mathbb{Z}_2 invariant ν can be generalized to the case of **three-dimensional topological insulators** which are characterized by four \mathbb{Z}_2 invariants $(\nu_0; \nu_1\nu_2\nu_3)$ (Refs. 7,8). The surface states of a three-dimensional topological insulator can be described in terms of a two-dimensional momentum \mathbf{k} within the surface Brillouin zone. Like in the two-dimensional case of a topological insulator—the quantum spin Hall insulator—there are four time-reversal invariant momenta Γ_a present in the surface Brillouin zone, where any state must be at least two-fold degenerate by virtue of Kramers’ theorem. These Kramers pairs of surface states form two-dimensional **Dirac points** in the surface band structure. Between any pair Γ_a and Γ_b of the special points in the surface Brillouin zone those surface states may connect

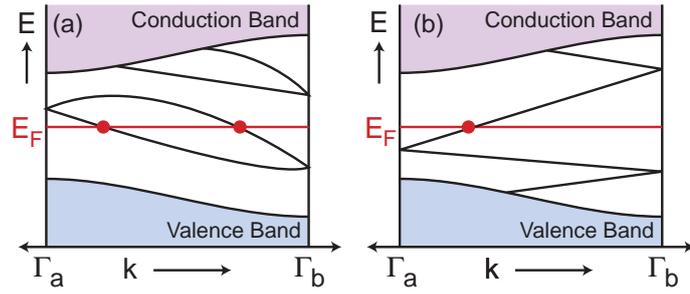


Figure 2.11. Electronic band structure between two time-reversal invariant momenta Γ_a and Γ_b as function of crystal momentum k . In panel (a) the surface states cross the Fermi energy E_F an even number of times, whereas in panel (b) the number of crossing points is odd. Note that an odd number of crossings leads to topologically protected metallic boundary states, as discussed in the main text. Figure taken from Ref. 28.

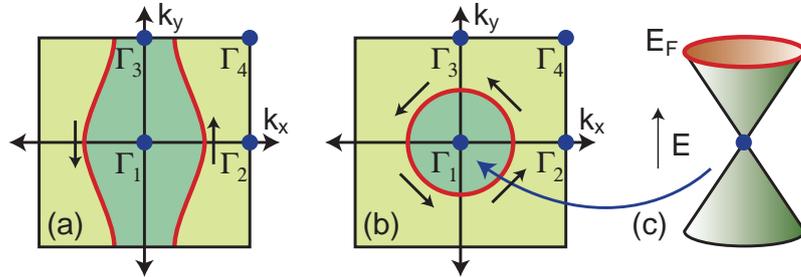


Figure 2.12. a, b Schematic picture of the Fermi surface (red) in the surface Brillouin zone for (a) a weak ($\nu_0 = 0$) and (b) a strong ($\nu_0 = 1$) topological insulator, respectively. The small arrows indicate the electron spin which is locked to the momentum k due to time-reversal symmetry. (c) In the simplest case of a strong topological insulator, the Fermi surface encircles a single Dirac point with linear dispersion relation. Figure taken from Ref. 28.

in such a way that they intersect the chemical potential either an even number of times [cf. Fig. 2.11 (a)] or an odd number of times [cf. Fig. 2.11 (b)]. Which of the two alternatives is realized is determined by the three \mathbb{Z}_2 invariants ν_1 , ν_2 , and ν_3 for the three independent surfaces of a crystal.

In the simplest case, a 3D topological insulator can be constructed by stacking layers of the 2D quantum spin Hall insulator on top of each other. As a result, the helical edge states of the layers hybridize and form surface states. Fig. 2.12 (a) shows one particular situation of a Fermi surface for weakly coupled layers stacked along the z direction, where the Fermi surface encloses an even number of time-reversal invariant momenta on the surface (Γ_1 and Γ_3 or equivalently Γ_2 and Γ_4). This state of matter is called a **weak topological insulator** with “trivial” topological index $\nu_0 = 0$. It has been shown that the other three indices ($\nu_1\nu_2\nu_3$) can be interpreted as Miller indices describing the orientation of the 2D quantum spin Hall layers (see, e.g., Ref. 8). Although surface states must be present for clean surfaces, they can be localized by disorder since time-reversal symmetry does not protect those states.

On the other hand, $\nu_0 = 1$ describes a **strong topological insulator** which cannot be interpreted in terms of the 2D quantum spin Hall insulator. In a strong topological insulator, the Fermi surface encircles an odd number of Kramers degenerate Dirac points on the 2D surface of the crystal [cf. Fig. 2.12 (b) and (c)]. In the simplest case of a single Dirac point, the 2D surface

metal can be described in terms of a Dirac Hamiltonian with linear spectrum:

$$\mathcal{H}_{\text{surf}}(\mathbf{k}) = v_F \mathbf{k} \cdot \boldsymbol{\sigma}. \quad (2.83)$$

Here, v_F is the Fermi velocity of the surface metal, and $\boldsymbol{\sigma}$ describes the electron's spin. As a result, the surface electronic band structure is similar to that of graphene, except that there is just a single Dirac fermion instead of four as in graphene (due to valley and spin degeneracies). Note that the electron spin is locked to the momentum \mathbf{k} due to time-reversal symmetry, leading to the notion of **helical surface states** in strong topological insulators. Moreover, an electron going around the Dirac point effectively rotates its spin by 2π , and therefore it picks up a non-trivial Berry phase of π . Note, however, that the description of the two-dimensional surface of a strong topological insulator in terms of a single Dirac fermion seemingly violates the **fermion doubling theorem** which states that for a time-reversal invariant system Dirac points must always come in pairs⁵⁰. The resolution to this puzzle is that the partner Dirac fermion resides on the opposite surface of the solid, restoring the counting scheme for Dirac fermions. Finally, the surface states of a strong topological insulator cannot be localized even for strong disorder as long as the bulk band gap remains intact⁵¹. Along with the presence of time-reversal invariance, this is often quoted as the major reason for the topological protection of the surface states against disorder and non-magnetic impurities.

2.5.3 Experimental realizations of strong topological insulators

In recent years, a class of 3D compounds—exemplified by Bi_2Te_3 , Bi_2Se_3 , and Sb_2Te_3 —were identified as strong topological insulators using ARPES (angle-resolved photoemission spectroscopy) measurements, where the surface state consists of a single Dirac cone^{52–54}.

From an experimental point of view, the topological insulators are challenging to realize, because there are two critical requirements that must be fulfilled simultaneously: First, the bulk of the sample must be free of magnetic impurities, because otherwise time-reversal symmetry is broken, and a gap for the surface states opens. Second, the Fermi energy must lie within the bulk gap. Chen *et al.* report in recent works on the realization of topological insulators in Bi_2Se_3 samples, in which they were able to precisely control the position of the Fermi energy^{52,55}. Moreover, they managed to introduce a controlled amount of magnetic dopants to break the time-reversal symmetry, which allows to study gapped surface Dirac fermions as well.

The experimental results of the electronic band structure of undoped Bi_2Se_3 obtained by ARPES measurements, shown in Fig. 2.13, beautifully exemplify the existence of the Dirac-like spectrum of the surface states in topological insulators. Similar to the band structure of Bi_2Te_3 (not shown here, cf. Ref. 54), aside from the Fermi surface of the surface-state band (indicated as SSB in Fig. 2.13), there is also the Fermi surface of the bulk conduction band (BCB). Note that the bottom of the bulk conduction band is located about 190 meV above the Dirac point labeled by E_d which is indicative of a direct bulk band gap [cf. Fig. 2.13 (c)]. As pointed out by Chen *et al.*, the Dirac point in Bi_2Se_3 is actually a better candidate than Bi_2Te_3 for realizing the topological surface Dirac fermion state, because in the latter system the surface Dirac point is below the top of the bulk valence band (BVB). However, experimentally one finds that the bulk conductivity is not small due to a large impurities concentration which makes it hard to detect the topological surface states⁵⁴.

Fig. 2.13 (b) shows a cross section plot of the band structure for various binding energies. Here, we observe how the band structure of the surface state evolves from the single Dirac point at the center of the surface Brillouin zone into a hexagonally shaped Fermi surface.

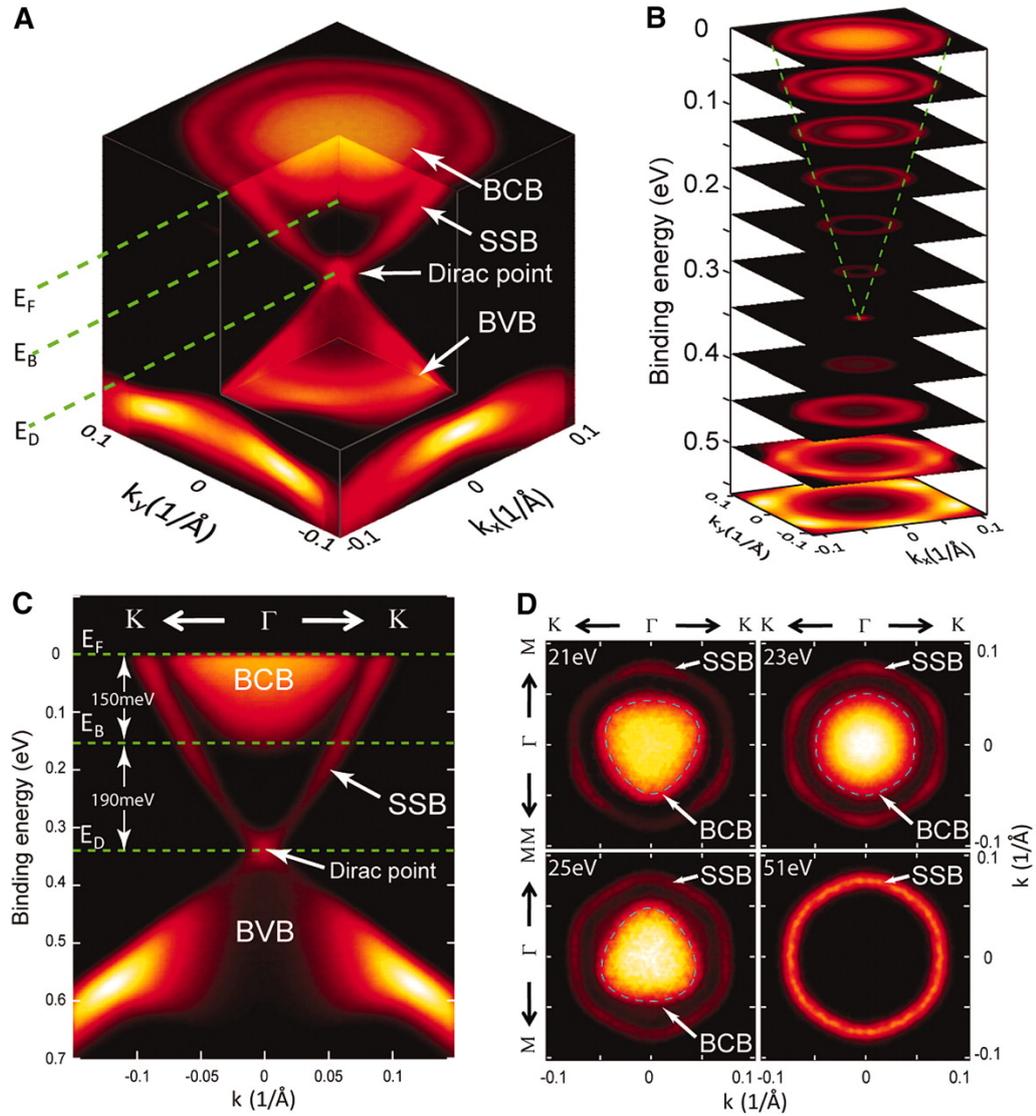


Figure 2.13. Experimental measurement of the electronic band structure of undoped Bi_2Se_3 measured by angle-resolved photo-emission spectroscopy (ARPES). (a) The bulk conduction band (BCB), bulk valence band (BVB), and surface state band (SSB) are indicated, along with the Fermi energy (E_F), the bottom of the BCB (E_B), and the Dirac point (E_D). (b) Constant-energy contours of the band structure show the evolution of the surface state band from the Dirac point to a hexagonal shape (indicated by green dashed lines). (c) Band structure along the $K - \Gamma - K$ direction, where Γ is the center of the hexagonal surface Brillouin zone, and the K and M points are the corner and the midpoint of the side of the Brillouin zone, respectively. The BCB bottom is about 190 meV above E_D and 150 meV below E_F . (d) Photon energy-dependent Fermi surface maps which have been symmetrized according to the crystal symmetry. Figure after Ref. 55.

Chen *et al.* also confirmed the nature of the hexagonal Fermi surface of the surface states by the photon energy-dependent ARPES, as shown in Fig. 2.13 (d). Since the shape of the Fermi surfaces does not vary upon changing the energy of the incident photons, the Fermi surface of the surface Dirac fermions is truly two-dimensional, whereas the Fermi surface of the bulk conduction band in the center clearly changes, which can be traced back to the k_z dependence of the dispersion relation.

Finally, topological insulators are also expected to occur in many other materials and not only in binary compounds like Bi_2Te_3 , Bi_2Se_3 , and Sb_2Te_3 . Promising candidates for topological insulators are, for example, certain iridium-based pyrochlore materials⁵⁶⁻⁵⁹, but also the family of ternary Heusler compounds^{60,61} are under current investigation.

CHAPTER 3

Electronic Band Structure of Topological Insulators

In condensed matter physics, the **electronic band structure** of a solid describes the allowed energies an electron may have in a periodic crystalline potential, thereby forming **energy bands** which are separated by **energy gaps** or **band gaps** from one another. In general, **band theory** describes the behavior of electrons in crystalline solids by assuming the existence of such energy bands, and it explains very well many physical properties such as optical absorption or electrical resistivity. From a theoretical vantage point, the description of electrons in a solid is generically a many-body problem and thus a complicated task. The full Hamiltonian of a perfect crystal does not only contain single-electron potentials describing the interactions of the electrons with the massive atomic nuclei, but also pair potentials describing the electron-electron interactions have to be taken into account. While many different approaches to the problem of calculating the band structure have been formulated in the past decades, in this thesis we only consider the so-called **tight-binding approach** which assumes that the electrons are tightly bound to the constituent atoms and can be described by a time-independent single-particle Schrödinger equation.

In a simplified picture, we can identify three major types of band structures describing **insulators**, **semiconductors**, and **metals**. Concerning the electronic properties, the only difference between an insulator and a conductor is that the band gap between the so-called **valence band** and the **conduction band** is much larger in the insulator than in the conductor. Here, the valence band describes the highest occupied band, in analogy to the valence electrons of individual atoms, whereas the lowest unoccupied band is referred to as the conduction band, because current can only flow when electrons are excited to the conduction band, for example by thermal fluctuations or by application of a gate voltage. The gap between valence band and conduction band strongly influences, among other things, the electrical and optical properties of a material, and is thus an important property of the solid.

In contrast to this simple picture, a **topological insulator** is a material which behaves like an insulator in the bulk of the sample, but exhibits metallic states on its surfaces which allow for charge transport or current. To be precise, the electronic band structure resembles an ordinary band insulator in the bulk of the sample, where the Fermi level sits between the valence band and the conduction band. However, on the surfaces of a topological insulator there exist states which (i) are exponentially localized at the surfaces, (ii) traverse the band gap, and (iii) allow for conduction on the surfaces. Importantly, those surface states are protected by certain fundamental symmetries of the system—time-reversal symmetry, particle-hole symme-

try, or chiral symmetry—and by the inherent topological properties of the bulk band structure (see, e.g., Refs. 62–67). Topologically protected boundary states were first predicted to occur in graphene nanoribbons^{3,4} and quantum well heterostructures of **mercury telluride** (HgTe) sandwiched between layers of **cadmium telluride** (CdTe) due to a particular **band inversion** in HgTe, and they were observed experimentally in those quantum wells in 2007^{5,6,68}. Later, topological insulators were also predicted on the surfaces of 3D bulk materials, forming a novel type of a two-dimensional electron gas, where the electron’s spin is locked to the momentum^{7,8}. The first experimentally realized 3D topological insulator was discovered in bismuth antimony (BiSb)^{69,70}, but soon after the topologically protected surface states were also observed in a number of materials such as pure antimony (Sb), bismuth selenide (Bi₂Se₃), bismuth telluride (Bi₂Te₃) and antimony telluride (Sb₂Te₃) using angle-resolved photoemission spectroscopy (ARPES)^{28,52,53,71,72}. Since then, many more materials have been studied and some of them are now believed to exhibit topological surface states as well^{60,61,73}.

In this chapter, we focus on the electronic properties and topological surface states of mercury telluride (HgTe) and cadmium telluride (CdTe). To develop a better understanding of the electronic band structure of both HgTe and CdTe, this chapter has been divided into several parts: First, we give a brief introduction into the general tight-binding approach to the electronic band structure of solids, following the textbooks by Yu and Cardona⁷⁴, Winkler⁷⁵, and Kittel⁷⁶. After that, we review the effect of spin-orbit interactions coupling the spin to the orbital angular momentum of the electron. Since the strength of the spin-orbit interaction scales with the atomic number of the atoms, the spin-orbit coupling is expected to be important for heavy elements such as mercury, cadmium, and telluride. Importantly, the energy correction of the levels due to spin-orbit interactions leads to an inversion of the bands in HgTe compared to CdTe which results in topologically protected, chiral edge channels, as will be discussed below. In the third section, we describe the crystal structure and the crystal symmetries of HgTe and CdTe, both showing a zinc-blende-type lattice. We then compute the bulk band structure and density of states of those materials using a 20-band tight-binding Hamiltonian which was already studied in the 1970’s and 1980’s^{77–81}. After that we focus on layered CdTe/HgTe/CdTe quantum well heterostructures which are synthetic structures containing a thin layer of HgTe sandwiched between two layers of CdTe with large band gap. In particular, we briefly review the theoretical aspects and the experimental discovery of the quantum spin Hall insulator in those heterostructures. In the last two sections, we first discuss the effect of biaxial strain on the band structure of HgTe which is generated by the epitaxial growth of HgTe on a CdTe substrate. Here, we show that biaxial strain opens up a direct band gap at the center of the Brillouin zone, but also an indirect band gap which we obtain from the density of states. In principle, this observation paves the way for topologically protected states on the surfaces of a strained 3D HgTe samples, because the surface states are no longer coupled to the metallic bulk states^{9,10}. In the last section, we construct a minimal model for 3D topological insulators which is based on the states of strained HgTe close to the Fermi level. We discuss in detail the electronic band structure and the phase diagram which shows both strong and weak topological insulator phases. In the next chapter, we will investigate the quantum Hall effect in this model and discuss the relation between the edge channels and the so-called θ -term¹⁹.

3.1 Tight-binding approach to the electronic band structure

In developing the **tight-binding approach** for electrons in a solid we start from the assumption that the electrons are tightly bound to the nuclei, just like in the atomic limit. Bringing the atoms closer together, the separation of the atoms becomes comparable to the lattice constant

in solids, so that their wave functions overlap, and it is possible to approximate the electronic wave functions in the solid by linear combinations of the atomic wave functions. Accordingly, this approach is also known as **linear combination of atomic orbitals** or **LCAO** approach⁸². The justification for these approximations is that in a covalently bonded semiconductor such as mercury telluride (HgTe) we can distinguish between two kinds of electronic states. On the one hand, electrons in the conduction bands are delocalized and can be approximated well by nearly free electrons. On the other hand, the valence band electrons are mainly localized in the bonds and behave more like atomic states, so that their wave functions should be very similar to **bonding orbitals** found in molecules.

To formulate the tight-binding approach, let us start by decomposing the position of an atom in the crystalline lattice into $\mathbf{r}_{jl} = \mathbf{R}_j + \mathbf{r}_l$, where \mathbf{R}_j denotes the position of the j th primitive cell of the Bravais lattice, and \mathbf{r}_l is the position of the l th atom within the unit cell. Furthermore, let us consider a crystal Hamiltonian \mathcal{H} which is given by the sum of the atomic Hamiltonians describing the atomic orbitals of the nuclei and an interaction term describing the electronic interaction (hopping) between different atoms. As mentioned above, we can construct electronic wave functions as linear combinations of the atomic wave functions if the interaction between the atoms is weak. Due to the translational invariance of the crystal, those wave functions can be written in terms of Bloch functions $\Phi_{ml\mathbf{k}}(\mathbf{r})$ as

$$\Phi_{ml\mathbf{k}}(\mathbf{r}) = \frac{1}{\sqrt{N}} \sum_j e^{i\mathbf{k} \cdot \mathbf{r}_{jl}} \phi_{ml}(\mathbf{r} - \mathbf{r}_{jl}), \quad (3.1)$$

where m labels the electronic state of the l th atom with crystal momentum \mathbf{k} , and N is the number of unit cells in the crystal. The atomic orbitals $\phi_{ml}(\mathbf{r} - \mathbf{r}_{jl})$, also known as **Löwdin orbitals**, are constructed from the usual atomic orbital wave functions in such a way that wave functions centered at different atoms are orthogonal to each other:

$$\int d^3r \phi_{ml}^*(\mathbf{r} - \mathbf{r}_{jl}) \phi_{m'l'}(\mathbf{r} - \mathbf{r}_{j'l'}) = \delta_{j,j'} \delta_{l,l'} \delta_{m,m'}. \quad (3.2)$$

As a consequence from this orthogonality relation, the electronic wave functions $\Phi_{ml\mathbf{k}}$ are also orthogonal to each other:

$$\int d^3r \Phi_{ml\mathbf{k}}^*(\mathbf{r}) \Phi_{m'l'\mathbf{k}'}(\mathbf{r}) = \delta(\mathbf{k} - \mathbf{k}') \delta_{l,l'} \delta_{m,m'}, \quad (3.3)$$

where $\delta(x)$ denotes the Dirac delta function. The eigenfunctions $\Psi_{\mathbf{k}}(\mathbf{r})$ of the Hamiltonian \mathcal{H} are superpositions of the Bloch functions $\Phi_{ml\mathbf{k}}(\mathbf{r})$:

$$\Psi_{\mathbf{k}}(\mathbf{r}) = \sum_{m,l} c_{ml}(\mathbf{k}) \Phi_{ml\mathbf{k}}(\mathbf{r}) \quad (3.4)$$

with complex coefficients $c_{ml}(\mathbf{k}) \in \mathbb{C}$ to be determined. To compute the set of eigenvalues $\mathcal{E}_{\mathbf{k}}$ and eigenfunctions $\Psi_{\mathbf{k}}(\mathbf{r})$ of the Hamiltonian \mathcal{H} , we multiply the stationary Schrödinger equation $\mathcal{H}\Psi_{\mathbf{k}}(\mathbf{r}) = \mathcal{E}_{\mathbf{k}}\Psi_{\mathbf{k}}(\mathbf{r})$ by the wave function $\Psi_{\mathbf{k}}^*(\mathbf{r})$ from the left and integrate over the spatial coordinate \mathbf{r} . Using the orthogonality of the electronic wave functions $\Phi_{ml\mathbf{k}}(\mathbf{r})$ we obtain a linear set of equations for the coefficients c_{ml} :

$$\sum_{m,l} c_{ml}(\mathbf{k}) (\mathcal{H}_{ml,m'l'}(\mathbf{k}) - \mathcal{E}_{\mathbf{k}} \delta_{l,l'} \delta_{m,m'}) \doteq 0. \quad (3.5)$$

Here, $\mathcal{H}_{ml,m'l'}(\mathbf{k})$ represents the matrix element $\langle \Phi_{ml\mathbf{k}} | \mathcal{H} | \Phi_{m'l'\mathbf{k}} \rangle$ which in terms of the Löwdin orbitals $\phi_{ml}(\mathbf{r})$ can be written as:

$$\mathcal{H}_{ml,m'l'}(\mathbf{k}) = \frac{1}{N} \sum_{j,j'} \exp[i\mathbf{k} \cdot (\mathbf{r}_{jl} - \mathbf{r}_{j'l'})] \langle \phi_{m'l'}(\mathbf{r} - \mathbf{r}_{jl}) | \mathcal{H} | \phi_{ml}(\mathbf{r} - \mathbf{r}_{j'l'}) \rangle. \quad (3.6)$$

For a material modeled by N bands including orbital and spin degrees of freedom the matrix elements $\mathcal{H}_{ml,m'l'}(\mathbf{k})$ form a Hermitian matrix of size $N \times N$ which can be easily diagonalized, thereby solving Eq. (3.5). As a consequence, the eigenfunctions $\Psi_{n\mathbf{k}}(\mathbf{r})$ and the eigenenergies $\mathcal{E}_{n\mathbf{k}}$ of the Hamiltonian \mathcal{H} are labeled by the so-called **band index**, which takes on the values $n = 1, \dots, N$, and the crystal momentum \mathbf{k} . A plot of the electron energies $\mathcal{E}_{n\mathbf{k}}$ as function of the crystal momentum \mathbf{k} is usually known as the **electronic band structure** of the crystal.

Assuming that the electrons are tightly bound to the atomic nuclei the overlap of orbital wave functions decreases very fast with increasing distance between the unit cells j and j' . Hence, instead of summing over all unit cells j and j' in the crystal, we can approximate the matrix element $\mathcal{H}_{ml,m'l'}(\mathbf{k})$ by restricting the sum only over the nearest neighbors of an atom. In that sense, j will be summed over the atom itself and its nearest neighbors:

$$\mathcal{H}_{ml,m'l'}(\mathbf{k}) = \sum_j \exp[i\mathbf{k} \cdot (\mathbf{R}_j + \mathbf{r}_l - \mathbf{r}_{l'})] \langle \phi_{ml}(\mathbf{r} - \mathbf{r}_{jl}) | \mathcal{H} | \phi_{m'l'}(\mathbf{r} - \mathbf{r}_{j'l'}) \rangle. \quad (3.7)$$

Of course, this approximation is not limited to nearest-neighbor interactions only, but one can easily include interactions between second-nearest or even further neighbors if needed by a straightforward generalization of the matrix elements. Additional matrix elements describing interactions over longer distances can be used, for example, to increase the accuracy of the tight-binding calculations.

To conclude, within the tight-binding approach (3.5), the band structure of electrons in a solid can be obtained by calculating the eigenvalues of a Bloch Hamiltonian $\mathcal{H}(\mathbf{k})$, which describes the interactions of electrons tightly bound to the nuclei, as function of the crystal momentum \mathbf{k} . The overlap of the orbital wave functions describes the hopping of electrons between different atoms and leads to the formation of delocalized conduction bands. The valence band electrons, on the other hand, are localized in the bonding orbitals and in general responsible for the elastic properties of the crystal which are not at the focus of this chapter.

3.2 Spin-orbit interactions

It is well known from atomic physics that the electron spin is coupled to the orbital angular momentum via the spin-orbit interaction which is a relativistic correction to the usual tight-binding Hamiltonian. The spin-orbit correction can be visualized when we shift from the standard frame of reference, where the electron orbits the nucleus, into a reference frame, where the electron is stationary and the nucleus instead orbits the electron. In that case, the orbiting nucleus generates an effective magnetic field which couples to the electron's spin. A Taylor expansion of the Dirac equation in powers of $(v/c)^2$, where c denotes the speed of light, leads to several terms which are responsible for the hyperfine splitting of the atomic energy levels. In particular, the atomic Hamiltonian for the **spin-orbit interaction** is given by

$$H_{\text{SO}} = \frac{\hbar}{4m^2c^2} (\nabla V(\mathbf{r}) \times \mathbf{p}) \cdot \boldsymbol{\sigma} = \frac{\hbar}{4m^2c^2} \epsilon_{ijk} \partial_j V(\mathbf{r}) p_k \sigma_i, \quad (3.8)$$

where $V(\mathbf{r})$ is the Coulomb potential of the atom with atomic number Z , and $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)^T$ is the usual vector of Pauli spin matrices [cf. Eq. (2.68)]. This energy correction is a factor $(Z\alpha)^2$

smaller than the Coulomb energy, with $\alpha \approx 1/137$ the fine-structure constant. However, since the Coulomb interaction $V(\mathbf{r})$ and thus the strength of the spin-orbit interaction both scale with the atomic number Z of the atom, we expect the spin-orbit coupling to be important only for materials which are made of heavy elements with large Z , such as CdTe and HgTe. Therefore, especially for those compounds we have to include spin-orbit interactions into the tight-binding Hamiltonian to properly describe their electronic band structures.

In atomic physics, the Hamiltonian for the spin-orbit interaction is often expressed in terms of the **orbital angular momentum operator** l and the **spin operator** s as

$$H_{\text{SO}} = 2\lambda_{\text{SO}} \mathbf{l} \cdot \mathbf{s}, \quad (3.9)$$

where λ_{SO} is usually referred to as the **spin-orbit coupling parameter**. In that representation it becomes apparent that there is no spin-orbit correction to the s -wave-like states which have an orbital angular momentum $l = 0$.

To develop a representation of the spin-orbit interaction (3.8) which can be used in the tight-binding approach, let us start with a brief review of some facts about orbital wave functions. From quantum mechanics we know that the orbital electronic wave functions $|l, m_l\rangle$ are classified according to their **orbital angular momentum** l , and they can be chosen as eigenstates of the z component \hat{l}_z of the orbital angular momentum operator \hat{l} . The eigenvalues of \hat{l}_z , usually denoted by m_l or just m if there is no ambiguity with other quantum numbers, are called **magnetic quantum numbers** and take the values $-l, -l+1, \dots, l-1, l$. Furthermore, the eigenfunctions of the spin-orbit interaction (3.9) are eigenstates of the **total angular momentum** $\hat{j} = \hat{l} + \hat{s}$ and its z component, \hat{j}_z . In general, the magnetic quantum number m_j corresponding to \hat{j}_z takes the values $-j, -j+1, \dots, j-1, j$. Since we consider a single electron ($s = \frac{1}{2}$) in a p orbital state ($l = 1$), the eigenvalues of \hat{j} can be either $j = l + s = \frac{3}{2}$ or $j = l - s = \frac{1}{2}$. The common eigenfunctions $|j, m_j, l\rangle$ of \hat{j} and \hat{j}_z can be expressed as superpositions of the eigenfunctions $|l, m_l, \sigma\rangle$ of the orbital angular momentum \hat{l} and spin \hat{s} . To be precise, we obtain a $j = \frac{3}{2}$ quartet:

$$|\frac{3}{2}, +\frac{3}{2}, 1\rangle = |1, 1, \uparrow\rangle, \quad (3.10a)$$

$$|\frac{3}{2}, +\frac{1}{2}, 1\rangle = \frac{1}{\sqrt{3}}(|1, 1, \downarrow\rangle + \sqrt{2}|1, 0, \uparrow\rangle), \quad (3.10b)$$

$$|\frac{3}{2}, -\frac{1}{2}, 1\rangle = \frac{1}{\sqrt{3}}(|1, -1, \uparrow\rangle + \sqrt{2}|1, 0, \downarrow\rangle), \quad (3.10c)$$

$$|\frac{3}{2}, -\frac{3}{2}, 1\rangle = |1, -1, \downarrow\rangle, \quad (3.10d)$$

and a $j = \frac{1}{2}$ doublet:

$$|\frac{1}{2}, +\frac{1}{2}, 1\rangle = \frac{1}{\sqrt{3}}(|1, 0, \uparrow\rangle - \sqrt{2}|1, 1, \downarrow\rangle), \quad (3.10e)$$

$$|\frac{1}{2}, -\frac{1}{2}, 1\rangle = \frac{1}{\sqrt{3}}(|1, 0, \downarrow\rangle - \sqrt{2}|1, -1, \uparrow\rangle). \quad (3.10f)$$

Using the relation

$$2\hat{l} \cdot \hat{s} = (\hat{l} + \hat{s})^2 - \hat{l}^2 - \hat{s}^2 = \hat{j}^2 - \hat{l}^2 - \hat{s}^2 = j(j+1) - l(l+1) - s(s+1), \quad (3.11)$$

we can easily calculate the matrix elements of the spin-orbit interaction (3.9). As a result, for

the p orbital states we obtain:

$$\langle \frac{3}{2}, m_j, 1 | 2\lambda_{\text{SO}} \hat{\mathbf{l}} \cdot \hat{\mathbf{s}} | \frac{3}{2}, m_j, 1 \rangle = \lambda_{\text{SO}}, \quad (3.12a)$$

$$\langle \frac{1}{2}, m_j, 1 | 2\lambda_{\text{SO}} \hat{\mathbf{l}} \cdot \hat{\mathbf{s}} | \frac{1}{2}, m_j, 1 \rangle = -2\lambda_{\text{SO}}. \quad (3.12b)$$

Thus, in the basis of total angular momentum states, the spin-orbit interaction is diagonal:

$$\mathcal{H}_{\text{SO}} = \lambda_{\text{SO}} \text{diag}(1, 1, 1, 1, -2, -2). \quad (3.13)$$

Obviously, the effect of the spin-orbit interaction is to split the $j = \frac{3}{2}$ states (3.10a–3.10d) from the $j = \frac{1}{2}$ states (3.10e–3.10f), and for the $j = \frac{3}{2}$ and $j = \frac{1}{2}$ states this **spin-orbit splitting** is equal to $\Delta_{\text{SO}} = 3\lambda_{\text{SO}}$.

As mentioned before, the spin-orbit interaction $H_{\text{SO}} = 2\lambda_{\text{SO}} \mathbf{l} \cdot \mathbf{s}$ couples the electron spin σ to the orbital angular momentum. Using the following basis of p orbital states,

$$\{|p, \sigma\rangle\} = \{|p_x, \uparrow\rangle, |p_y, \uparrow\rangle, |p_z, \uparrow\rangle, |p_x, \downarrow\rangle, |p_y, \downarrow\rangle, |p_z, \downarrow\rangle\}, \quad (3.14)$$

this coupling between orbital angular momentum and electron spin is easy to observe in the following 6×6 Hamiltonian for the spin-orbit interaction:

$$\mathcal{H}_{\text{SO}} = \lambda_{\text{SO}} \begin{bmatrix} 0 & i & 0 & 0 & 0 & 1 \\ -i & 0 & 0 & 0 & 0 & i \\ 0 & 0 & 0 & -1 & -i & 0 \\ 0 & 0 & -1 & 0 & -i & 0 \\ 0 & 0 & i & i & 0 & 0 \\ 1 & -i & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (3.15)$$

Apparently, the off-diagonal matrix elements couple the spin-up and spin-down states of the different p orbitals and thus lead to a non-trivial mixing of states. In particular, since the spin-orbit interaction is quite strong in heavy elements like Hg and Cd, this coupling may lead to sizable and important corrections of the electronic band structure, and thus spin-orbit interactions cannot be neglected, but have to be considered within in the tight-binding approach.

In the following, we will develop a suitable Bloch Hamiltonian for the outermost, partially filled s and p orbitals of HgTe and CdTe. Note that in HgTe the spin-orbit interactions are so strong that they result in a band inversion within the electronic band structure. This is one of the major reasons for the appearance of topologically protected edge and surface states in two-dimensional quantum wells and strained 3D HgTe samples, as discussed later.

3.3 Electronic band structure of HgTe and CdTe

In this thesis we are mainly interested in the electronic band structures of **mercury telluride** (HgTe) and **cadmium telluride** (CdTe). Both HgTe and CdTe are binary II-VI compounds, where cadmium (Cd) and mercury (Hg) are group-II elements in the periodic table of elements with atomic numbers $Z = 48$ and $Z = 80$, respectively, while tellurium (Te) is a group-VI element with atomic number $Z = 52$. Furthermore, both CdTe and HgTe are **tetrahedrally bonded** semiconductors in which each atom is **four-fold coordinated**, *i.e.*, each atom is surrounded by four nearest-neighbor atoms forming a tetrahedron (see Fig. 3.1). The ionicity of the covalent bonds in II-VI compounds increases the Coulomb interaction between the ions, leading to an increased band gap in the electronic band structure. CdTe, for example, has a direct band gap

of about 1.6 eV at room temperature, but HgTe is actually a **zero-band gap semiconductor** or **semimetal**. Due to the strong spin-orbit coupling in HgTe, the bands close to the Fermi level are inverted compared to CdTe, as will be discussed in detail in section 3.3.4. As mentioned before, the band inversion is one of the reasons for the edge states in quantum well heterostructures made of CdTe and HgTe which is discussed in section 3.4. In this section, however, we first discuss the crystal properties of bulk HgTe and CdTe, and then construct a 20-band tight-binding Hamiltonian following Refs. 77–81 which describes both the covalent bonding of the outermost, partially filled sp^3 orbital states as well as some excited conduction band states. After that we review the electronic band structure of those materials and introduce the terminology frequently used in semiconductor physics in section 3.3.4.

3.3.1 The crystal structure and symmetries of HgTe and CdTe

The crystal structure of HgTe and CdTe is the same as in the zinc-blende lattice with a two-atomic unit cell consisting of anionic telluride (Te^{2-}) and cationic mercury (Hg^{2+}) or cationic cadmium (Cd^{2+}). The zinc-blende lattice shown in Fig. 3.2 (a) is a **face-centered cubic (fcc)** Bravais lattice [see Fig. 3.2 (b)], and its **primitive lattice vectors** are given by

$$\mathbf{a}_1 = \frac{a}{2} (0, 1, 1)^T, \quad \mathbf{a}_2 = \frac{a}{2} (1, 0, 1)^T, \quad \mathbf{a}_3 = \frac{a}{2} (1, 1, 0)^T, \quad (3.16)$$

where a is the length of the crystallographic unit cell, and \mathbf{v}^T denotes the transpose of the vector \mathbf{v} , as before. The position of each atom in the crystalline solid can be decomposed into $\mathbf{r}_{jl} = \mathbf{R}_j + \mathbf{r}_l$, where \mathbf{R}_j denotes the position of the j th unit cell of the Bravais lattice, and \mathbf{r}_l is the position of the l th atom within the unit cell. Note that the mercury telluride lattice has two atoms per unit cell, *i.e.*, $l = 1, 2$ only, and those atoms are located at $\mathbf{r}_1 = (0, 0, 0)^T$ and $\mathbf{r}_2 = \frac{a}{4}(1, 1, 1)^T$. Hence, the positions of the nearest-neighbor atoms in the zinc-blende lattice are given by

$$\begin{aligned} \mathbf{d}_1 &= \frac{a}{4} (1, 1, 1)^T, & \mathbf{d}_2 &= \frac{a}{4} (1, -1, -1)^T, \\ \mathbf{d}_3 &= \frac{a}{4} (-1, 1, -1)^T, & \mathbf{d}_4 &= \frac{a}{4} (-1, -1, 1)^T. \end{aligned} \quad (3.17)$$

The reciprocal lattice is defined in terms of three **primitive reciprocal lattice vectors** $\mathbf{b}_1, \mathbf{b}_2$, and \mathbf{b}_3 which are related to the direct lattice vectors $\mathbf{a}_1, \mathbf{a}_2$, and \mathbf{a}_3 by $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$, where

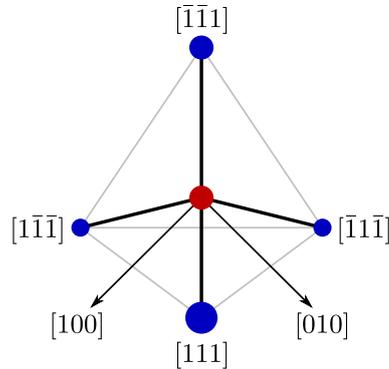


Figure 3.1. A tetrahedron with a mercury atom (red filled circle) in the center surrounded by four telluride atoms (blue filled circles). The thick black lines indicate the bonds in the high-symmetry directions $[111]$, $[1\bar{1}\bar{1}]$, $[\bar{1}1\bar{1}]$, and $[\bar{1}\bar{1}1]$. The coordinate axes are shown in black, where the $[001]$ axis points out of the plane.

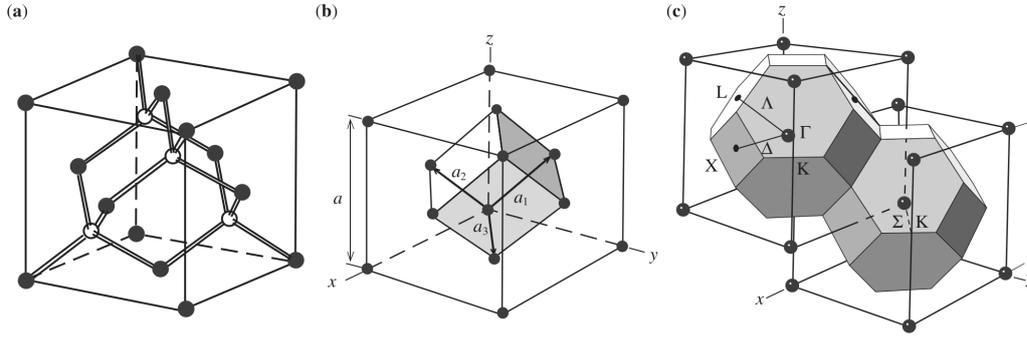


Figure 3.2. (a) The crystal structure of the mercury telluride (HgTe) lattice. (b) The face-centered cubic lattice with a set of primitive lattice vectors \mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3 . (c) The reciprocal lattice of the fcc lattice and the first Brillouin zone. Points of high symmetry are denoted by $\Gamma = (0, 0, 0)^T$, $X = \frac{2\pi}{a}(1, 0, 0)^T$, $K = \frac{2\pi}{a}(\frac{3}{4}, \frac{3}{4}, 0)^T$, and $L = \frac{2\pi}{a}(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})^T$, while the three high-symmetry directions $[100]$, $[110]$, and $[111]$ in the Brillouin zone are denoted by Δ , Σ and Λ . Figure after Ref. 74.

δ_{ij} is the Kronecker delta. In three dimensions, a general solution for this relation is given by

$$\mathbf{b}_i = \epsilon_{ijk} \pi \frac{\mathbf{a}_j \times \mathbf{a}_k}{|(\mathbf{a}_1 \times \mathbf{a}_2) \cdot \mathbf{a}_3|},$$

where i, j , and k represent a cyclic permutation of the three indices 1, 2 and 3, ϵ_{ijk} denotes the fully anti-symmetric tensor with $\epsilon_{123} = 1$, and $|(\mathbf{a}_1 \times \mathbf{a}_2) \cdot \mathbf{a}_3|$ is the volume of the primitive cell. The reciprocal lattice points of the fcc lattice form the **body-centered cubic** (bcc) lattice, so that the primitive reciprocal lattice vectors of the HgTe crystal are given by

$$\mathbf{b}_1 = \frac{2\pi}{a}(-1, 1, 1)^T, \quad \mathbf{b}_2 = \frac{2\pi}{a}(1, -1, 1)^T, \quad \mathbf{b}_3 = \frac{2\pi}{a}(1, 1, -1)^T. \quad (3.18)$$

The first **Brillouin zone** of the fcc lattice, *i.e.*, the unit cell of the reciprocal lattice, is also shown in Fig. 3.2 (c). Since the reciprocal lattice vectors are obtained from the lattice vectors \mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3 , the symmetries of the Brillouin zone are determined by the symmetries of the crystal lattice, as discussed below.

The **space group** of the zinc-blende structure has the number 216 and is denoted by T_d^2 in Schönflies notation or $F\bar{4}3m$ in international notation, where the symbol $F\bar{4}3m$ describes a face-centered cubic lattice with four-fold improper rotations about the $\langle 001 \rangle$ axes, three-fold rotations about the $\langle 111 \rangle$ axes, and reflections on the $\{110\}$ mirror planes. Note that this space group is **symmorphic**, *i.e.*, apart from the lattice translations, all generating symmetry operations leave one common point fixed which implies that generators of such a space group can only include the point-group operations, *i.e.*, rotations, reflections, inversions, and improper rotations. The corresponding **point group** of the zinc-blende lattice has 24 elements which are identical to the elements of the point group of a tetrahedron, denoted by T_d . The point group symmetry operations of the zinc-blende crystal are defined with respect to the three mutually perpendicular crystallographic axes, where the origin is placed at one of the two atoms in the unit cell. With this choice of coordinates, the 24 symmetry operations are listed in Table 3.1. The elements of this point group can be divided into five classes $\{E\}$, $\{8C_3\}$, $\{3C_2\}$, $\{6S_4\}$, and $\{6\sigma\}$, because rotations by the same angle with respect to equivalent axes belong to one class, while all reflections on equivalent planes belong to a different class. Hence, T_d has five **irreducible representations** which are usually denoted by A_1 , A_2 , E , T_1 , and T_2 , where E denotes the identity operation in group theory.

class	symmetries of the space group #216 / T_d^2 / $F\bar{4}3m$
$\{E\}$	identity
$\{8C_3\}$	three-fold clockwise and counterclockwise rotations of 120° about the $[111]$, $[\bar{1}\bar{1}\bar{1}]$, $[1\bar{1}\bar{1}]$, and $[\bar{1}\bar{1}1]$ axes
$\{3C_2\}$	two-fold rotations of 180° about each of the $[100]$, $[010]$, and $[001]$ axes
$\{6S_4\}$	four-fold improper clockwise and counterclockwise rotations of 90° about the $[100]$, $[010]$ and $[001]$ axes
$\{6\sigma\}$	reflections on each of the (110) , $(\bar{1}\bar{1}0)$, (101) , $(10\bar{1})$, (011) , and $(01\bar{1})$ planes

Table 3.1. The space group T_d^2 of the zinc-blende lattice has the same symmetry operators as the point group of the tetrahedron, denoted by T_d .

molecular notation	Koster notation	BSW notation
A_1	Γ_1	Γ_1
A_2	Γ_2	Γ_2
E	Γ_3	Γ_{12}
T_2	Γ_4	Γ_{15}
T_1	Γ_5	Γ_{25}

Table 3.2. List of commonly used notations for the irreducible representations of the T_d point group. Note that Γ_4 and Γ_5 are sometimes interchanged in the literature.

However, using the notation A_1 , A_2 , E , T_1 , and T_2 to label the irreducible representation of the T_d group is more commonly used in molecular physics than in semiconductor physics, where a different notation is used. First, note that the wave functions of a crystal with wave vector \mathbf{k} at the center of the Brillouin zone, *i.e.*, the Γ point, transform in a way that is specified by the irreducible representations of the point group of the crystal. Thus, the corresponding Bloch functions at the Γ point can also be classified according to these irreducible representations. Historically, in semiconductor physics it is more common to use “ Γ ” plus a subscript “ i ” to label the irreducible representations of T_d , but there are two different conventions used for labeling the same irreducible representation by the index i . The first convention to label the five irreducible representations by $\Gamma_1, \dots, \Gamma_5$ is due to Koster and is used most commonly in the literature, but the Bouckaert-Smoluchewski-Wigner (BSW) notation ($\Gamma_1, \Gamma_2, \Gamma_{12}, \Gamma_{15}, \Gamma_{25}$) is also sometimes used. For reference, the different notations for the T_d point group are listed in Table 3.2.

3.3.2 Tight-binding Hamiltonian for HgTe and CdTe

To properly describe the sp^3 hybridized bonds of the those materials, we formulate the tight-binding Hamiltonian $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_{SO}$ for CdTe and HgTe in terms of an sp^3s^* basis which describes the covalent bonding of the outermost, partially filled sp^3 orbital states on the one hand and some excited s^* conduction band states on the other hand⁷⁷⁻⁸¹. To be specific, we consider the following basis of 20 orbital angular momentum states, $\{|\nu, \alpha, \sigma, \mathbf{k}\rangle\}$, where $\nu \in \{a, c\}$ denotes the atom and distinguishes between the anion ($\nu = a$) and the cation ($\nu = c$) in the unit cell. $\alpha \in \{s, p_x, p_y, p_z, s^*\}$ denotes the orbital symmetry, $\sigma \in \{\uparrow, \downarrow\}$ is the electron spin, and \mathbf{k} denotes the crystal momentum. This basis set of states has the distinct advantage that

one can in a straightforward and direct way write down the tight-binding Hamiltonian \mathcal{H}_0 describing the hopping of electrons between orbitals on nearest-neighbor atoms. On the other hand, as discussed before, the spin-orbit Hamiltonian \mathcal{H}_{SO} then contains off-diagonal matrix elements coupling different spin states with angular momentum $l = 1$ which makes the analytical calculation of the band structure much more complicated. As a consequence, for generic crystal momentum \mathbf{k} we have to numerically compute the electronic band structure by means of exact diagonalization. Note that one can also write down the tight-binding Hamiltonian in the basis of total angular momentum states, $\{|\nu, j, m_j, l, \mathbf{k}\rangle\}$, in which the spin-orbit Hamiltonian is diagonal, but the corresponding tight-binding Hamiltonian becomes more complicated. We refer the reader to Ref. 79 for an explicit form of the 20-band tight-binding Hamiltonian in terms of those eigenstates.

The 20×20 Bloch Hamiltonian matrix \mathcal{H}_0 describing the hopping processes of electrons between nearest-neighbor atoms turns out to be block-diagonal in momentum space and can conveniently be written in terms of the 5×5 matrices $\mathfrak{h}_{\nu\nu'}^0(\mathbf{k})$ as follows:

$$\mathcal{H}_0(\mathbf{k}) = \begin{bmatrix} \mathfrak{h}_{aa}^0(\mathbf{k}) & 0 & \mathfrak{h}_{ac}^0(\mathbf{k}) & 0 \\ 0 & \mathfrak{h}_{aa}^0(\mathbf{k}) & 0 & \mathfrak{h}_{ac}^0(\mathbf{k}) \\ \mathfrak{h}_{ca}^0(\mathbf{k}) & 0 & \mathfrak{h}_{cc}^0(\mathbf{k}) & 0 \\ 0 & \mathfrak{h}_{ca}^0(\mathbf{k}) & 0 & \mathfrak{h}_{cc}^0(\mathbf{k}) \end{bmatrix}. \quad (3.19)$$

First, note that the diagonal block matrices $\mathfrak{h}_{\nu\nu}^0(\mathbf{k})$ describe the energies of the s , p , and s^* orbital states of the anion ($\nu = a$) or the cation ($\nu = c$) with respect to each other. Hence, the block matrices $\mathfrak{h}_{\nu\nu}^0(\mathbf{k})$ are diagonal themselves and independent of the crystal momentum \mathbf{k} :

$$\mathfrak{h}_{\nu\nu}^0(\mathbf{k}) = \begin{bmatrix} E_{s\nu} & 0 & 0 & 0 & 0 \\ 0 & E_{p\nu} & 0 & 0 & 0 \\ 0 & 0 & E_{p\nu} & 0 & 0 \\ 0 & 0 & 0 & E_{p\nu} & 0 \\ 0 & 0 & 0 & 0 & E_{s^*\nu} \end{bmatrix}. \quad (3.20)$$

On the other hand, the off-diagonal block matrices $\mathfrak{h}_{ac}^0(\mathbf{k})$ and $\mathfrak{h}_{ca}^0(\mathbf{k})$ describe the hopping of electrons between the different orbital states of anion and cation, independent of the spin orientation. To be specific, $\mathfrak{h}_{ac}^0(\mathbf{k})$ takes the following form:

$$\mathfrak{h}_{ac}^0(\mathbf{k}) = \begin{bmatrix} V_{ss} g_0(\mathbf{k}) & V_{sp} g_1(\mathbf{k}) & V_{sp} g_2(\mathbf{k}) & V_{sp} g_3(\mathbf{k}) & 0 \\ V_{ps} g_1(\mathbf{k}) & V_{xx} g_0(\mathbf{k}) & V_{xy} g_3(\mathbf{k}) & V_{xy} g_2(\mathbf{k}) & V_{ps^*} g_1(\mathbf{k}) \\ V_{ps} g_2(\mathbf{k}) & V_{xy} g_3(\mathbf{k}) & V_{xx} g_0(\mathbf{k}) & V_{xy} g_1(\mathbf{k}) & V_{ps^*} g_2(\mathbf{k}) \\ V_{ps} g_3(\mathbf{k}) & V_{xy} g_2(\mathbf{k}) & V_{xy} g_1(\mathbf{k}) & V_{xx} g_0(\mathbf{k}) & V_{ps^*} g_3(\mathbf{k}) \\ 0 & V_{s^*p} g_1(\mathbf{k}) & V_{s^*p} g_2(\mathbf{k}) & V_{s^*p} g_3(\mathbf{k}) & 0 \end{bmatrix}. \quad (3.21)$$

For example, the matrix element of two s orbital states on the anion and cation is given by:

$$\langle a, s, \sigma, \mathbf{k} | \mathcal{H}_0 | c, s, \sigma, \mathbf{k} \rangle = \langle s | \mathcal{H}_0 | s \rangle \sum_{j=1}^4 e^{i\mathbf{k} \cdot \mathbf{d}_j} \equiv V_{ss} g_0(\mathbf{k}), \quad (3.22)$$

where we have taken the anion ($\nu = a$) to be located at the origin, and \mathbf{d}_j ($j = 1, \dots, 4$) denotes the positions of its four nearest-neighbor cations ($\nu = c$) in the zinc-blende crystal. The other matrix elements can also be expressed in terms of the four phase factors $\exp(i\mathbf{k} \cdot \mathbf{d}_j)$ and the overlap parameters $V_{ss\sigma}$, $V_{sp\sigma}$, $V_{ps\sigma}$, $V_{pp\sigma}$, $V_{pp\pi}$, $V_{s^*p\sigma}$, and $V_{ps^*\sigma}$. Hopping from the

excited s^* orbital to the s and s^* orbitals of neighboring sites have been neglected for simplicity, following Refs. 79,80. In appendix A we have summarized the general construction scheme of (i) the overlap parameters and (ii) the above matrix elements. However, note that the p orbital states have to be decomposed into the so-called σ and π components due to the crystal structure of the zinc-blende lattice, as shown in appendix A in detail. As a consequence, this decomposition introduces a factor of $\cos\theta = \pm 1/\sqrt{3}$, where the sign depends on the relative orientation of the positive and negative lobes of the p orbitals. It is thus convenient to define the following **overlap parameters** for the tunneling or hopping of electrons between different orbitals:

$$\begin{aligned} V_{ss} &\equiv V_{ss\sigma}, & V_{sp} &\equiv \frac{V_{sp\sigma}}{\sqrt{3}}, & V_{ps} &\equiv \frac{V_{ps\sigma}}{\sqrt{3}}, \\ V_{xx} &\equiv \frac{V_{pp\sigma} + 2V_{pp\pi}}{3}, & V_{xy} &\equiv \frac{V_{pp\sigma} - V_{pp\pi}}{3}, \\ V_{s^*p} &\equiv \frac{V_{sp^*\sigma}}{\sqrt{3}}, & V_{ps^*} &\equiv \frac{V_{ps^*\sigma}}{\sqrt{3}}. \end{aligned} \quad (3.23)$$

and the relative orientation of the anions and cations in the zinc-blende lattice enters in the functions $g_j(\mathbf{k})$ with $j = 0, \dots, 3$. In contrast to the construction of overlap parameters for molecules shown in appendix A, we allow that, for example, V_{sp} and V_{ps} differ from each other, so that one can use an additional parameter in fitting the calculated band structure to first-principle band structure calculations or experiments. The four momentum-dependent functions $g_j(\mathbf{k})$ (with $j = 0, \dots, 3$) obtained from summing over the phase factors $\exp(i\mathbf{k} \cdot \mathbf{d}_j)$ are then given by

$$g_0(\mathbf{k}) \equiv \frac{1}{4} \sum_{j=1}^4 e^{i\mathbf{k} \cdot \mathbf{d}_j} = \frac{1}{4} (e^{i\mathbf{k} \cdot \mathbf{d}_1} + e^{i\mathbf{k} \cdot \mathbf{d}_2} + e^{i\mathbf{k} \cdot \mathbf{d}_3} + e^{i\mathbf{k} \cdot \mathbf{d}_4}), \quad (3.24a)$$

$$g_1(\mathbf{k}) \equiv \frac{1}{4} \sum_{j=1}^4 \text{sgn}(\mathbf{d}_j \cdot \hat{\mathbf{x}}) e^{i\mathbf{k} \cdot \mathbf{d}_j} = \frac{1}{4} (e^{i\mathbf{k} \cdot \mathbf{d}_1} + e^{i\mathbf{k} \cdot \mathbf{d}_2} - e^{i\mathbf{k} \cdot \mathbf{d}_3} - e^{i\mathbf{k} \cdot \mathbf{d}_4}), \quad (3.24b)$$

$$g_2(\mathbf{k}) \equiv \frac{1}{4} \sum_{j=1}^4 \text{sgn}(\mathbf{d}_j \cdot \hat{\mathbf{y}}) e^{i\mathbf{k} \cdot \mathbf{d}_j} = \frac{1}{4} (e^{i\mathbf{k} \cdot \mathbf{d}_1} - e^{i\mathbf{k} \cdot \mathbf{d}_2} + e^{i\mathbf{k} \cdot \mathbf{d}_3} - e^{i\mathbf{k} \cdot \mathbf{d}_4}), \quad (3.24c)$$

$$g_3(\mathbf{k}) \equiv \frac{1}{4} \sum_{j=1}^4 \text{sgn}(\mathbf{d}_j \cdot \hat{\mathbf{z}}) e^{i\mathbf{k} \cdot \mathbf{d}_j} = \frac{1}{4} (e^{i\mathbf{k} \cdot \mathbf{d}_1} - e^{i\mathbf{k} \cdot \mathbf{d}_2} - e^{i\mathbf{k} \cdot \mathbf{d}_3} + e^{i\mathbf{k} \cdot \mathbf{d}_4}). \quad (3.24d)$$

Finally, the spin-orbit interaction contains off-diagonal matrix elements coupling spin-up and spin-down electronic states of different p orbital states as discussed previously. In the basis of orbital angular momentum states we obtain:

$$\mathcal{H}_{\text{SO}} = \begin{bmatrix} \mathfrak{h}_{aa}^{\text{SO}}(\uparrow\uparrow) & \mathfrak{h}_{aa}^{\text{SO}}(\uparrow\downarrow) & 0 & 0 \\ \mathfrak{h}_{aa}^{\text{SO}}(\downarrow\uparrow) & \mathfrak{h}_{aa}^{\text{SO}}(\downarrow\downarrow) & 0 & 0 \\ 0 & 0 & \mathfrak{h}_{cc}^{\text{SO}}(\uparrow\uparrow) & \mathfrak{h}_{cc}^{\text{SO}}(\uparrow\downarrow) \\ 0 & 0 & \mathfrak{h}_{cc}^{\text{SO}}(\downarrow\uparrow) & \mathfrak{h}_{cc}^{\text{SO}}(\downarrow\downarrow) \end{bmatrix} \quad (3.25)$$

with the 5×5 matrices $h_{\nu\nu}^{\text{SO}}(\sigma\sigma')$ given by

$$\mathfrak{h}_{\nu\nu}^{\text{SO}}(\uparrow\uparrow) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -i\lambda_\nu & 0 & 0 \\ 0 & i\lambda_\nu & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \mathfrak{h}_{\nu\nu}^{\text{SO}}(\uparrow\downarrow) = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda_\nu & 0 \\ 0 & 0 & 0 & -i\lambda_\nu & 0 \\ 0 & -\lambda_\nu & i\lambda_\nu & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (3.26)$$

In addition, the spin-orbit interaction for spin-up and spin-down states is related by

$$\mathfrak{h}_{\nu\nu}^{\text{SO}}(\downarrow\downarrow) = [\mathfrak{h}_{\nu\nu}^{\text{SO}}(\uparrow\uparrow)]^*, \quad \mathfrak{h}_{\nu\nu}^{\text{SO}}(\downarrow\uparrow) = [\mathfrak{h}_{\nu\nu}^{\text{SO}}(\uparrow\downarrow)]^\dagger, \quad (3.27)$$

where \mathfrak{h}^* and \mathfrak{h}^\dagger denote the complex conjugate and the Hermitian conjugate of the matrix \mathfrak{h} , respectively.

3.3.2.1 Energy levels at the Γ point

To give an explicit example of how to compute the electronic band structure of the HgTe and CdTe and to develop a better understanding of the spectrum at the center of the Brillouin zone, let us calculate the eigenenergies of the Hamiltonian matrix at the Γ point. For $\mathbf{k} = (0, 0, 0)^T$ we find with Eq. (3.24) that $g_0(\mathbf{k}) = 1$ and $g_j(\mathbf{k}) = 0$ for $j = 1, 2, 3$. As a result, the 20×20 tight-binding matrix simplifies into two 2×2 matrices for the bonding and anti-bonding s electrons and two 2×2 matrices describing the excited s^* orbitals:

$$\begin{bmatrix} E_{sa} & V_{ss} \\ V_{ss} & E_{sc} \end{bmatrix}, \quad \begin{bmatrix} E_{s^*a} & 0 \\ 0 & E_{s^*c} \end{bmatrix}. \quad (3.28a)$$

Moreover, the p orbital states are coupled due to spin-orbit interactions, resulting in two equivalent 6×6 matrices:

$$\begin{bmatrix} E_{pa} & -i\lambda_a & \lambda_a & V_{xx} & 0 & 0 \\ i\lambda_a & E_{pa} & -i\lambda_a & 0 & V_{xx} & 0 \\ \lambda_a & i\lambda_a & E_{pa} & 0 & 0 & 4V_{xx} \\ V_{xx} & 0 & 0 & E_{pc} & -i\lambda_c & \lambda_c \\ 0 & V_{xx} & 0 & i\lambda_c & E_{pc} & -i\lambda_c \\ 0 & 0 & V_{xx} & \lambda_c & i\lambda_c & E_{pc} \end{bmatrix}. \quad (3.28b)$$

These three matrices can be easily diagonalized, giving the following spectrum:

$$\mathcal{E}_{s\pm}(\Gamma) = \frac{E_{sa} + E_{sc}}{2} \pm \frac{1}{2} \sqrt{(E_{sa} - E_{sc})^2 + 4V_{ss}^2}, \quad (3.29a)$$

$$\mathcal{E}_{s^*\pm}(\Gamma) = \frac{E_{s^*a} + E_{s^*c}}{2} \pm \frac{E_{s^*a} - E_{s^*c}}{2}, \quad (3.29b)$$

$$\mathcal{E}_{p\pm}(\Gamma) = \frac{E_{pa} + E_{pc} + \lambda_a + \lambda_c}{2} \pm \frac{1}{2} \sqrt{(E_{pa} - E_{pc} + \lambda_a - \lambda_c)^2 + 4V_{xx}^2}, \quad (3.29c)$$

$$\mathcal{E}_{p\pm}(\Gamma) = \frac{E_{pa} + E_{pc} - 2\lambda_a - 2\lambda_c}{2} \pm \frac{1}{2} \sqrt{(E_{pa} - E_{pc} + 2\lambda_a - 2\lambda_c)^2 + 4V_{xx}^2}. \quad (3.29d)$$

In the atomic limit, *i.e.*, for vanishing inter-atomic interactions, $V_{ss} = V_{xx} = 0$, we observe the aforementioned **spin-orbit splitting** between the $j = \frac{3}{2}$ and $j = \frac{1}{2}$ states:

$$\Delta_{\text{SO}} = \begin{cases} 3\lambda_a & \text{for the anion } (\nu = a) \\ 3\lambda_c & \text{for the cation } (\nu = c) \end{cases} \quad (3.30)$$

with the corresponding magnitudes for anion and cation in the two-atomic unit cell. Note that this energy splitting of the $j = \frac{3}{2}$ and $j = \frac{1}{2}$ states at the Γ point is used to determine the strength of the spin-orbit coupling parameter λ_{SO} by fitting some experimental data to band structure calculations with and without spin-orbit interactions⁷⁹.

3.3.3 Bulk band structure and density of states of HgTe and CdTe

The empirical tight-binding parameters for the electronic band structures which are listed in Table 3.3 were obtained by Kobayashi, Sankey, and Dow by fitting numerical band structures of HgTe and CdTe at the Γ point and the X points of the Brillouin zone [see Fig. 3.2 (c)] to experimental data⁷⁹. Furthermore, the spin-orbit coupling parameters λ_a and λ_c were fitted from the spin-orbit splitting at the Γ point between the $j = \frac{3}{2}$ and $j = \frac{1}{2}$ states.

Figures 3.3 and 3.4 show the computed electronic band structure of CdTe and HgTe along the principal symmetry directions of the zinc-blende-type solid (left panels) and the corresponding density of states (right panels). For clarity, the bulk bands with so-called Γ_6 , Γ_7 , and Γ_8 symmetries are shown in red, green, and blue colors, respectively. These bands are particularly important in the CdTe/HgTe/CdTe quantum well heterostructures hosting two-dimensional variants of a topological insulator, which will be discussed in the following section. For example, in ordinary insulators like CdTe, spin-orbit interaction is not very strong, so that the Γ_8 bands are below the Γ_6 band (see Fig. 3.3). However, since Hg has a larger atomic number, $Z = 80$, the spin-orbit interactions are much larger in such heavy-element compounds which modifies the topology of the electronic band structure. In particular, the spin-orbit interaction leads to the aforementioned **band inversion** of the Γ_6 and Γ_8 bands and makes HgTe actually a semimetal. Also, note that the larger splitting of the peaks in the density of states of HgTe compared to CdTe is indicative of strong split-orbit interactions.

	CdTe	HgTe
on-site energies:		
E_{sa}	-8.891	-9.776
E_{pa}	0.915	0.155
E_{s^*a}	7.000	6.000
E_{sc}	-0.589	-1.404
E_{pc}	4.315	4.300
E_{s^*c}	7.500	6.500
overlap parameters:		
$4V_{ss}$	-4.779	-3.267
$4V_{sp}$	1.739	2.412
$4V_{ps}$	-4.767	-3.243
$4V_{xx}$	2.355	1.443
$4V_{xy}$	4.124	3.639
$4V_{s^*p}$	1.949	3.520
$4V_{ps^*}$	-2.649	-0.323
spin-orbit couplings:		
λ_a	0.367	0.299
λ_c	0.013	0.286

Table 3.3. Tight-binding parameters in electronvolt of the 20-band model for CdTe and HgTe including nearest-neighbor interactions only, obtained by Kobayashi, Sankey, and Dow⁷⁹. The overlap parameters V_{ss} , etc. are related to those in the standard notation due to Slater and Koster⁸² as follows: $V_{ss} = V_{ss\sigma}$, $V_{sp} = V_{sp\sigma}/\sqrt{3}$, $V_{ps} = -V_{ps\sigma}/\sqrt{3}$, $V_{xx} = (V_{pp\sigma} + 2V_{pp\pi})/3$, $V_{xy} = (V_{pp\sigma} - V_{pp\pi})/3$, $V_{s^*p} = V_{s^*p\sigma}/\sqrt{3}$, and $V_{ps^*} = -V_{ps^*\sigma}/\sqrt{3}$.

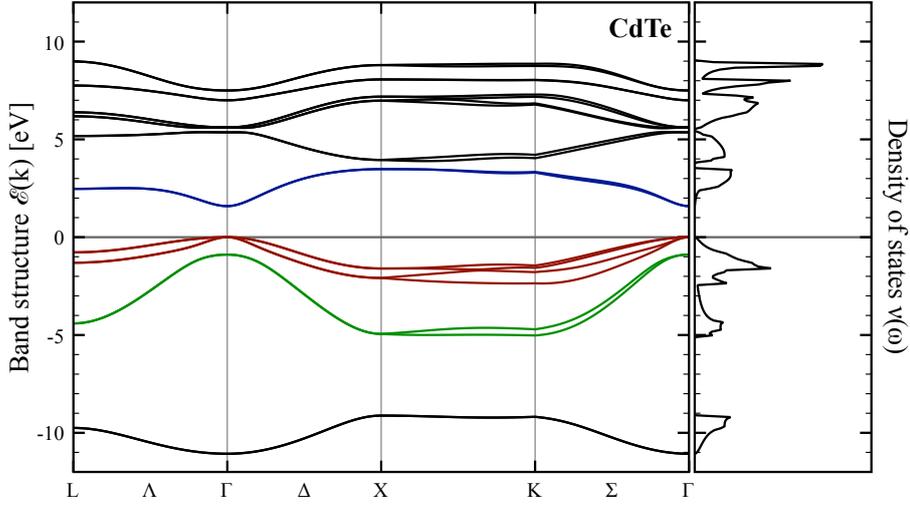


Figure 3.3. (a) Band structure $\mathcal{E}(\mathbf{k})$ of cadmium telluride (CdTe) as function of crystal momentum \mathbf{k} along the principal symmetry directions. The high-symmetry points are denoted by $L = \frac{2\pi}{a}(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})^T$, $\Gamma = (0, 0, 0)^T$, $X = \frac{2\pi}{a}(1, 0, 0)^T$, and $K = \frac{2\pi}{a}(\frac{3}{4}, \frac{3}{4}, 0)^T$, while the three high-symmetry directions [100], [110], and [111] in the Brillouin zone are denoted by Δ , Σ and Λ . The so-called Γ_6 and Γ_8 valence and conduction bands are shown in blue and red color, respectively, while the spin-orbit split-off Γ_7 valence band is indicated by green. (b) Density of states $\nu(\omega)$ of CdTe calculated numerically from the band structure $\mathcal{E}(\mathbf{k})$ performing exact diagonalization of the tight-binding Hamiltonian $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_{SO}$ for $5 \cdot 10^9$ random sampling points in momentum space.

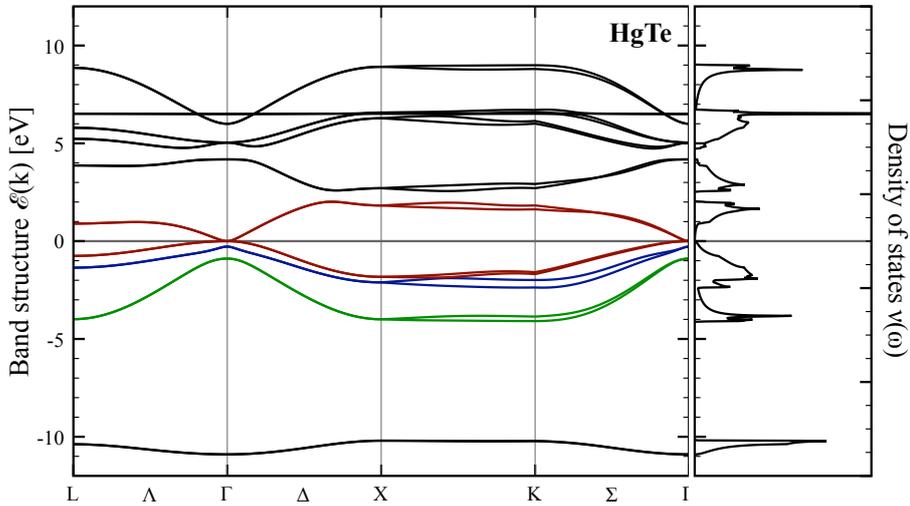


Figure 3.4. (a) Band structure $\mathcal{E}(\mathbf{k})$ of mercury telluride (HgTe) as function of crystal momentum \mathbf{k} along the principal symmetry directions. As in Fig. 3.3, the points of high-symmetry are denoted by L, Γ, X , and K , while the lines of high-symmetry in the [100], [110], and [111] directions are denoted by Δ, Σ and Λ . Similarly, the Γ_6 and Γ_8 bands are shown in blue and red color, respectively, while the Γ_7 valence band has a green color. Importantly, note that the Γ_6 and Γ_8 bands are inverted with respect to the usual band order as in CdTe. (b) Density of states $\nu(\omega)$ for HgTe calculated numerically from the band structure $\mathcal{E}(\mathbf{k})$ performing exact diagonalization of the tight-binding Hamiltonian $\mathcal{H} = \mathcal{H}_0 + \mathcal{H}_{SO}$ for $5 \cdot 10^9$ random sampling points in momentum space.

3.3.4 Irreducible representations, symmetries, and terminology of the bulk bands

In this section, we briefly introduce a certain notation of the electronic band structure of the diamond- and zinc-blende-type semiconductors which is based on group theory⁷⁴. Since the electrons move in the presence of a periodic crystal potential, their wave functions can be symmetrized to reflect the crystal symmetry, *i.e.*, the wave functions can be written in a form such that they belong to certain **irreducible representations** of the space group of the crystal. In particular, the symmetries of the wave functions close to the Γ point, *i.e.*, the center of the Brillouin zone are important, because the energy bands have a simple parabolic form in the vicinity of that point. In the following, we therefore introduce the terminology which is frequently used in the context of zinc-blende-type materials such as HgTe and CdTe, and we refer the reader to Ref. 74 for details.

Although we have introduced the Bloch Hamiltonian in terms of orbital angular momentum states labeled as $\{|\nu, \alpha, \sigma, \mathbf{k}\rangle\}$, where $\alpha \in \{s, p_x, p_y, p_z\}$ specifies the orbital symmetry of the wave function and $\sigma \in \{\uparrow, \downarrow\}$ the electron spin, we focus on a different basis of total angular momentum states which will be denoted by $\{|\nu, j, m_j, l, \mathbf{k}\rangle\}$ due to the strong spin-orbit interactions in heavy elements. Here, j is the total angular momentum, m_j the corresponding magnetic quantum number, and l the orbital angular momentum. From quantum mechanics we recall that the transformation between the two basis sets of wave functions can be performed in a straightforward manner:

$$|j, m_j, l\rangle = |l \pm \frac{1}{2}, m_j, l\rangle \equiv \alpha_{\pm} |l, m_j - \frac{1}{2}, \uparrow\rangle + \beta_{\pm} |l, m_j + \frac{1}{2}, \downarrow\rangle, \quad (3.31)$$

where the so-called **Clebsch-Gordan coefficients** for a spin 1/2 electron with angular momentum l are given by

$$\alpha_{\pm} \equiv \sqrt{\frac{l \pm m_j + \frac{1}{2}}{2l + 1}} \equiv \pm \beta_{\mp}, \quad \alpha_{+}^2 + \alpha_{-}^2 = \beta_{+}^2 + \beta_{-}^2 = 1. \quad (3.32)$$

In the literature, it is common to classify total angular momentum eigenstates according to their behavior under symmetry transformations of the crystal by means of group theory. Aside from the orbital part of the wave function, one has to take the spin of the electron into account as well. Groups containing symmetry operations of wave functions including the spin degree of freedom are known as **double groups**¹. In most cases it is sufficient to know only the **irreducible representations** for the double group at the center of the Brillouin zone, *i.e.*, the Γ point. In the case of the zinc-blende crystal, the 48 elements of the double group are divided into eight classes which are labeled by $\Gamma_1, \dots, \Gamma_8$, whose representations can be found, for example, in the textbook by Yu and Cardona⁷⁴. A careful inspection of the symmetries then reveals that the two-fold degenerate, *s*-like $|j = \frac{1}{2}, m_j = \pm \frac{1}{2}, l = 0\rangle$ states belong to the so-called **Γ_6 representation**⁷⁴. In addition, the two-fold degenerate, *p*-like $|j = \frac{1}{2}, m_j = \pm \frac{1}{2}, l = 1\rangle$ states belong to **Γ_7 representation**, whereas the four-fold degenerate, *p*-like $|j = \frac{3}{2}, m_j = \pm \frac{3}{2}, l = 1\rangle$ and $|j = \frac{3}{2}, m_j = \pm \frac{1}{2}, l = 1\rangle$ states belong to the **Γ_8 representation**, because this is the only four-dimensional representation in this group. Although those representations describe the symmetries of the corresponding atomic wave functions, the electronic bands in a solid inherit their symmetry properties and are labeled accordingly as **Γ_6 band**, **Γ_7 band**, and **Γ_8 band**, respectively.

¹It is beyond the scope of this thesis to describe double groups in detail. The reader should refer to the textbooks by Yu and Cardona⁷⁴ and Lew Yan Voon and Willatzen⁸³ and references therein for details on the definition of double groups.

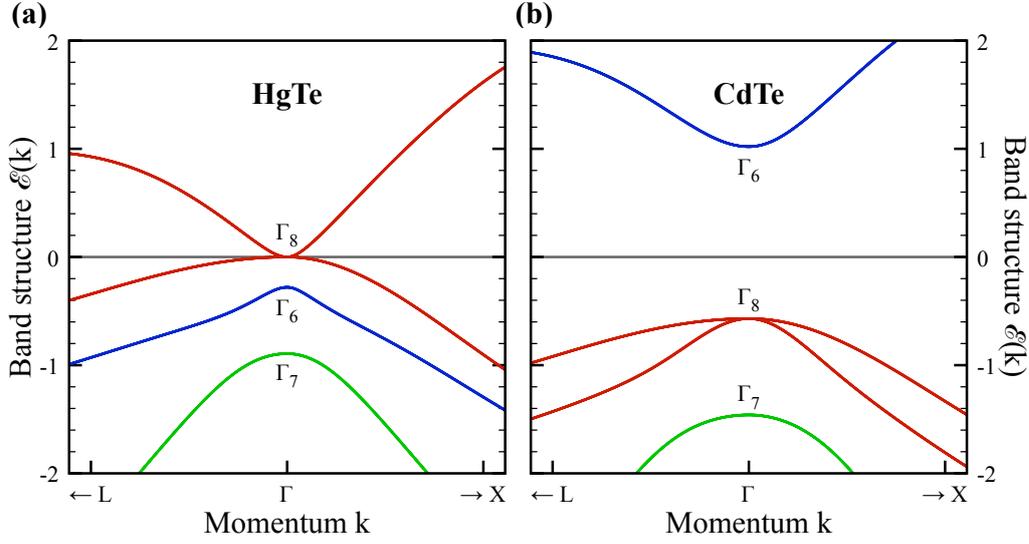


Figure 3.5. (a, b) Schematic picture of the bulk band structure of HgTe and CdTe close to the Fermi level at $E = 0$ eV. At the Γ point there are four important bands: The Γ_6 band is two-fold degenerate, stems from the $|j = \frac{1}{2}, m_j = \pm\frac{1}{2}, l = 0\rangle$ states, and shows an s -orbital like behavior. Moreover, the Γ_6 band is usually referred to as electron (E) band. The Γ_7 band is also two-fold degenerate, but stems from $|j = \frac{1}{2}, m_j = \pm\frac{1}{2}, l = 1\rangle$ states, and thus shows a p -orbital like behavior. Due to the spin-orbit splitting between the Γ_7 and Γ_8 bands, the Γ_7 band is usually called the spin-orbit split-off (SO) band. The Γ_8 band is, however, four-fold degenerate, because it originates from the $|j = \frac{3}{2}, m_j = \pm\frac{3}{2}, l = 1\rangle$ and $|j = \frac{3}{2}, m_j = \pm\frac{1}{2}, l = 1\rangle$ states with p -orbital like character. The $|j = \frac{3}{2}, m_j = \pm\frac{3}{2}, l = 1\rangle$ states are known as heavy-hole (HH) bands, while the $|j = \frac{3}{2}, m_j = \pm\frac{1}{2}, l = 1\rangle$ states are referred to as light-hole (LH) bands due to their different effective masses. Figure after Ref. 27.

group theory	$\mathbf{k} \cdot \mathbf{p}$ theory	$ j, m_j, l\rangle$ states	$ \alpha, \sigma\rangle$ states
$ \Gamma_6, +\frac{1}{2}\rangle$	$ E, +\frac{1}{2}\rangle$	$ \frac{1}{2}, +\frac{1}{2}, 0\rangle$	$ s, \uparrow\rangle$
$ \Gamma_6, -\frac{1}{2}\rangle$	$ E, -\frac{1}{2}\rangle$	$ \frac{1}{2}, -\frac{1}{2}, 0\rangle$	$ s, \downarrow\rangle$
$ \Gamma_7, +\frac{1}{2}\rangle$	$ SO, +\frac{1}{2}\rangle$	$ \frac{1}{2}, +\frac{1}{2}, 1\rangle$	$\frac{1}{\sqrt{3}} (p_x, \downarrow\rangle + i p_y, \downarrow\rangle + p_z, \uparrow\rangle)$
$ \Gamma_7, -\frac{1}{2}\rangle$	$ SO, -\frac{1}{2}\rangle$	$ \frac{1}{2}, -\frac{1}{2}, 1\rangle$	$\frac{1}{\sqrt{3}} (p_x, \uparrow\rangle - i p_y, \uparrow\rangle - p_z, \downarrow\rangle)$
$ \Gamma_8, +\frac{3}{2}\rangle$	$ HH, +\frac{3}{2}\rangle$	$ \frac{3}{2}, +\frac{3}{2}, 1\rangle$	$\frac{1}{\sqrt{2}} (p_x, \uparrow\rangle + i p_y, \uparrow\rangle)$
$ \Gamma_8, +\frac{1}{2}\rangle$	$ LH, +\frac{1}{2}\rangle$	$ \frac{3}{2}, +\frac{1}{2}, 1\rangle$	$\frac{1}{\sqrt{6}} (p_x, \downarrow\rangle + i p_y, \downarrow\rangle - 2 p_z, \uparrow\rangle)$
$ \Gamma_8, -\frac{1}{2}\rangle$	$ LH, -\frac{1}{2}\rangle$	$ \frac{3}{2}, -\frac{1}{2}, 1\rangle$	$\frac{-1}{\sqrt{6}} (p_x, \uparrow\rangle - i p_y, \uparrow\rangle + 2 p_z, \downarrow\rangle)$
$ \Gamma_8, -\frac{3}{2}\rangle$	$ HH, -\frac{3}{2}\rangle$	$ \frac{3}{2}, -\frac{3}{2}, 1\rangle$	$\frac{-1}{\sqrt{2}} (p_x, \downarrow\rangle + i p_y, \downarrow\rangle)$

Table 3.4. List of commonly used notations for the band topology of semiconductors like CdTe and HgTe. Note that the notations are often used synonymously in the literature.

There exists, however, another frequently used naming convention for the topology of the band structure which is related to the so-called $\mathbf{k} \cdot \mathbf{p}$ theory and describes the particular curvature of the bands at the Γ point in the Brillouin zone. In normal semiconductors and insulators, the Γ_6 band describes an *s*-like **electron band** (E), because it has positive curvature at the Γ point with effective mass $m_{\text{eff}} > 0$, *i.e.*, $\mathcal{E}_{\Gamma_6}(\mathbf{k}) \approx \mathbf{k}^2/(2m_{\text{eff}})$ [see Fig. 3.5 (b)]. However, the Γ_8 bands describing the *p*-like valence bands have negative effective masses $m_{\text{eff}} < 0$ describing hole-like excitations. Moreover, the $m_j = \pm\frac{3}{2}$ states are referred to as **heavy-hole band** (HH), while the $m_j = \pm\frac{1}{2}$ states are called the **light-hole band** (LH), because $m_{\text{eff},LH} < m_{\text{eff},HH}$. For reference, the different notations for the band topology are listed in Table 3.4.

As mentioned before, the Γ_7 and Γ_8 states with orbital angular momentum $l = 1$ are split by the spin-orbit interaction, $H_{\text{SO}} = 2\lambda_{\text{SO}} \hat{\mathbf{l}} \cdot \hat{\mathbf{s}}$ (see section 3.2). Typically, the magnitude of the spin-orbit splitting $\Delta_{\text{SO}} = 3\lambda_{\text{SO}}$ in a semiconductor with relatively light atoms is small and thus negligible. However, in semiconductors containing heavier elements such as CdTe or HgTe, the spin-orbit gap can be as large as the band gap, $\Delta_{\text{SO}} \approx 1$ eV, pushing down the Γ_7 states in energy (see Fig. 3.5). Consequently, the Γ_7 states are referred to as the **spin-orbit split-off band** (SO) in the literature and usually not taken into account. Moreover, in HgTe the energy correction of the Γ_8 bands due to spin-orbit interactions is so large that the Γ_8 bands are pushed above the Γ_6 band (see Fig. 3.5). This reversal in the band ordering of the Γ_8 and Γ_6 bands is one of the major reasons for the topological properties of HgTe.

3.4 CdTe/HgTe/CdTe quantum well heterostructures

In the context of topological insulators, we now discuss CdTe/HgTe/CdTe **quantum wells** which are synthetic heterostructures containing a thin layer of HgTe of thickness d_{HgTe} sandwiched between two thin layers of CdTe of equal thicknesses d_{CdTe} (see Fig. 3.6). Over the past two decades, fabrication techniques for quantum well heterostructures have greatly improved, and ultimately these achievements have led to the experimental discovery of the so-called **quantum spin Hall effect** and topological edge states in CdTe/HgTe/CdTe quantum wells⁵. These quantum well heterostructures show the interesting property of having either a normal or an inverted band structure, depending on the thickness of the sandwiched HgTe layer. Heuristically, in a large quantum well, the confinement of the electronic states within the quantum well is small, so we may expect that the electronic band structure of the quantum well at the Γ point mostly looks like HgTe which has an inverted band structure. However, upon reducing the thickness of the HgTe epilayer the confinement energy increases, thus shifting the energy bands. Eventually, the thickness of the HgTe epilayer will fall below a critical thickness, $d_c \approx 63$ Å, and the energy bands will start to align in a normal way, *i.e.*, the quantum well behavior is dominated by properties of CdTe (see Fig. 3.7). At the critical thickness d_c , the band inversion takes place, and one-dimensional helical edge states appear at the boundaries of the two-dimensional quantum well heterostructure, giving rise to the **quantum spin Hall effect** in the absence of any external magnetic field^{5,6,68}. In this section, we review the basic properties of CdTe/HgTe/CdTe quantum well heterostructures, and afterwards we study the electronic band structure of these quantum wells, depending on both growth direction and quantum well width.

One of the important properties of CdTe/HgTe/CdTe quantum wells is that the lattice constants of HgTe and CdTe are almost identical to each other:

$$a_{\text{HgTe}} = 6.46 \text{ \AA} \quad \text{and} \quad a_{\text{CdTe}} = 6.48 \text{ \AA}. \quad (3.33)$$

This nearly perfect lattice alignment allows for **pseudomorphic growth** and is desirable for

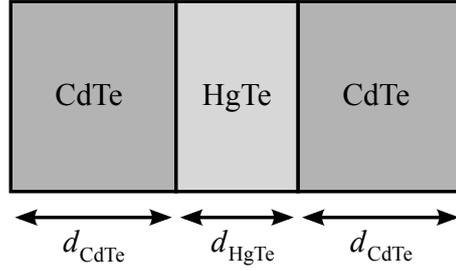


Figure 3.6. Schematic picture of a CdTe/HgTe/CdTe quantum well heterostructure in which the central HgTe epilayer of width d_{HgTe} is surrounded by CdTe epilayers of width d_{CdTe} on both sides.

high-quality quantum well heterostructures, because it allows to significantly reduce the number of crystal defects such as dislocations or vacancies. However, having the same lattice constants is not a necessary condition for pseudomorphic growth of the so-called **epitaxial layer** (epilayer) on the **substrate**. Growing the epilayer on a substrate with a different bulk lattice constant results in a strained, but otherwise perfect epilayer. Nevertheless, there is a limit to the thickness of a strained layer one can grow while maintaining a perfect lattice. Since the strain energy increases with the thickness, the epilayer may lower its total energy by relieving the strain beyond some **critical layer thickness** through the creation of dislocations. Obviously, the critical thickness for two materials grown on top of each other depends on the difference of their lattice constants. For CdTe/HgTe/CdTe heterostructures, the critical thickness for the HgTe epilayer is about 200 nm^{10} . Hence, to leading order, we neglect the lattice mismatch for narrow quantum wells ($d_{\text{HgTe}} \ll 200 \text{ nm}$) and use the bulk material properties of HgTe and CdTe to model the quantum wells. Later, in section 3.5 we consider strained HgTe systems and show that the application of biaxial strain leads to the opening of a small band gap at the Γ point.

Furthermore, due to the band difference between CdTe and HgTe, their conduction and valence band edges do not align with each other. The creation of quite sharp interfaces between CdTe and HgTe, however, allows to control the shape of this band gap discontinuity, and comparisons between experimental results and theoretical calculations have shown that these band edge discontinuities can be rather abrupt, making a simple square well a good approximation for the confinement potential in most quantum well structures. The difference between the band edges is also known as the **valence band offset**. For example, in semiconducting devices this band offset produces the potential which is responsible for confining the charge carriers (electrons or holes) in one layer only. For CdTe/HgTe/CdTe quantum wells, the valence band offset has been investigated in detail by a number of different methods^{84–89}, yielding approximately

$$E_{\text{VBO}} = (560 \pm 50) \text{ meV}. \quad (3.34)$$

In the following, we consider symmetric CdTe/HgTe/CdTe quantum well heterostructures which have a fixed number $N_{\text{CdTe}} = 8$ of CdTe layers surrounding a variable number N_{HgTe} of HgTe layers on each side. To calculate the electronic band structure of such a quantum well, we first have to construct a Bloch Hamiltonian matrix for a system with open boundary conditions in the growth direction of the quantum well and periodic boundary conditions in the two directions perpendicular to it. However, within the tight-binding approach these two-dimensional surfaces will always generate some surface states, even for trivial band insulators such as CdTe. These surface states correspond to dangling sp^3 -hybridized bonds at each surface atom, and thus strongly depend on the details of the surface. On the other hand, the topology of the conduction and valence bands of CdTe and HgTe are quite different, so that we may distinguish be-

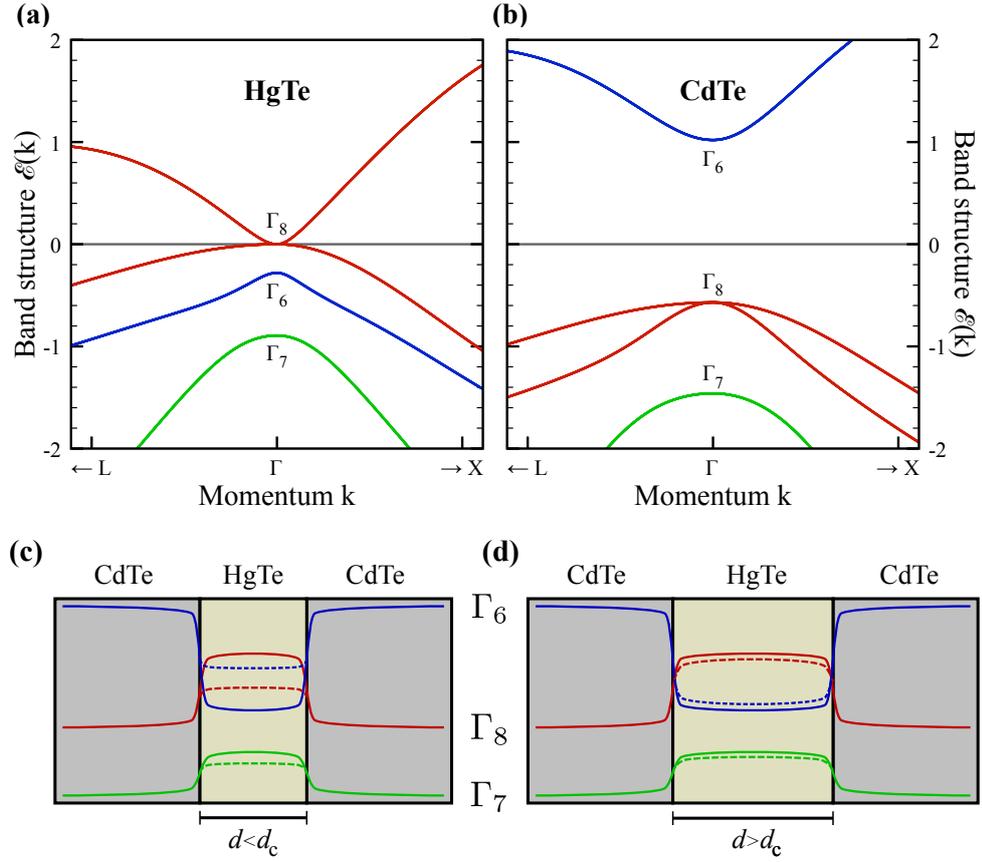


Figure 3.7. (a, b) Schematic picture of the bulk band structure of HgTe and CdTe close to the Fermi level at $E = 0$ eV. The important bands at the Γ point are given by the s orbital band Γ_6 and the p orbital bands Γ_7 and Γ_8 corresponding to the $j = \frac{1}{2}$ and $j = \frac{3}{2}$ total angular momentum states, respectively. (c, d) Sketch of the band structure in CdTe/HgTe/CdTe quantum wells as function of the quantum well width d . Solid lines indicate the bulk bands, and the dashed lines show the energy bands within the quantum well which differ from the bulk bands by some confinement energy [cf. Eq. (3.37)]. For $d < d_c$, the quantum well bands are dominated by the CdTe layers, while for $d > d_c$ the bands are inverted due to the strong spin-orbit interaction in HgTe. Figure after Ref. 27.

tween them. Since the Γ_6 , Γ_7 , and Γ_8 bands of CdTe are in the usual, non-inverted order, CdTe can be adiabatically connected to the vacuum by taking the band gap of CdTe to infinity, so that the topological properties of the interfaces between CdTe and HgTe are completely determined by HgTe. It turns out that the topological properties of the HgTe layer are insensitive to the details of the surface and do not depend on the exact surface renormalization scheme, so we may choose a surface regularization which removes the trivial surface states. Thus, in the following we consider periodic boundary conditions between the top and bottom CdTe layer, connecting the sp^3 -hybridized dangling bonds and thereby removing those trivial surface states.

To construct a tight-binding Bloch Hamiltonian for CdTe/HgTe/CdTe quantum well heterostructures we have to describe (i) the intra-layer hopping within a single epilayer and (ii) the inter-layer hopping between the CdTe and HgTe epilayers. Therefore we start by developing a tight-binding Bloch Hamiltonian for CdTe and HgTe in a **slab geometry** with planar surfaces describing the intra-layer hopping of electrons. After that we will couple those epilayers in the perpendicular direction, resulting in the tight-binding Hamiltonian for a quantum well heterostructure.

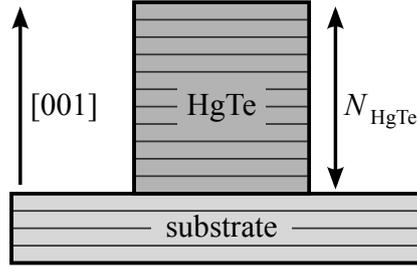


Figure 3.8. N_{HgTe} layers of HgTe are grown epitaxially along the [001] direction, forming a so-called slab geometry with planar surfaces whose normal vectors are along the [001] direction.

3.4.1 HgTe and CdTe in a (001) slab geometry

So far we have considered the properties of electrons in the bulk, *i.e.*, either in an infinite solid or a solid with periodic boundary conditions. Now let us assume that the crystal is finite in the [001] direction and that there are two impenetrable barriers (infinite energy barriers) associated with the two (001) surfaces which are separated by a distance d . These surfaces reflect the bulk Bloch wave functions along the [001] direction and break translational symmetry in that direction which has to be taken into account in the tight-binding approach to quantum well heterostructures. Furthermore, those surfaces reflecting the wave functions lead to standing waves whose wavelength λ takes on the discrete values

$$\lambda_n = \frac{2d}{n} \quad \text{with} \quad n = 1, 2, 3, \dots \quad (3.35)$$

In other words, the confinement leads to a quantization of states. For a free particle with mass m , the allowed wave vectors of the Bloch waves are given by

$$k_{z,n} = \frac{2\pi}{\lambda_n} = \frac{n\pi}{d} \quad \text{with} \quad n = 1, 2, 3, \dots \quad (3.36)$$

The confinement of the particle within a distance d along the z direction also increases the kinetic energy relative to the free case by the **confinement energy** which for $n = 1$ is given by

$$\mathcal{E}_{\text{conf}} = \frac{\hbar^2 k_{z,1}^2}{2m} = \frac{\hbar^2 \pi^2}{2m d^2}. \quad (3.37)$$

Note that this confinement energy is inversely proportional to the layer thickness, *i.e.*, $\mathcal{E}_{\text{conf}} \propto d^{-2}$, and thus goes to zero as d grows. Recall from quantum mechanics that in an infinite one-dimensional square well potential the energies of excited states are given by $\mathcal{E}_n = n^2 \mathcal{E}_{\text{conf}}$ with $n = 1, 2, 3, \dots$. The effect of this confinement on the electronic band structure is shown schematically in Fig. 3.7 (c, d). For small quantum wells, *i.e.*, small thickness d , the energy correction due to the confinement energy $\mathcal{E}_{\text{conf}}$ is large and leads to a sizable correction of the energy levels within the quantum well. In the opposite limit, *i.e.*, for large quantum wells, the confinement energy $\mathcal{E}_{\text{conf}}$ is rather small, and the HgTe epilayer in the center of the quantum well shows a band structure which is very close to the one of bulk HgTe.

Now, let us construct the tight-binding Hamiltonian for a single layer of either CdTe or HgTe. Since the translational symmetry is broken in the [001] direction, the construction of electronic wave functions can only include Bloch wave functions within the (001) plane perpendicular to the [001] direction.

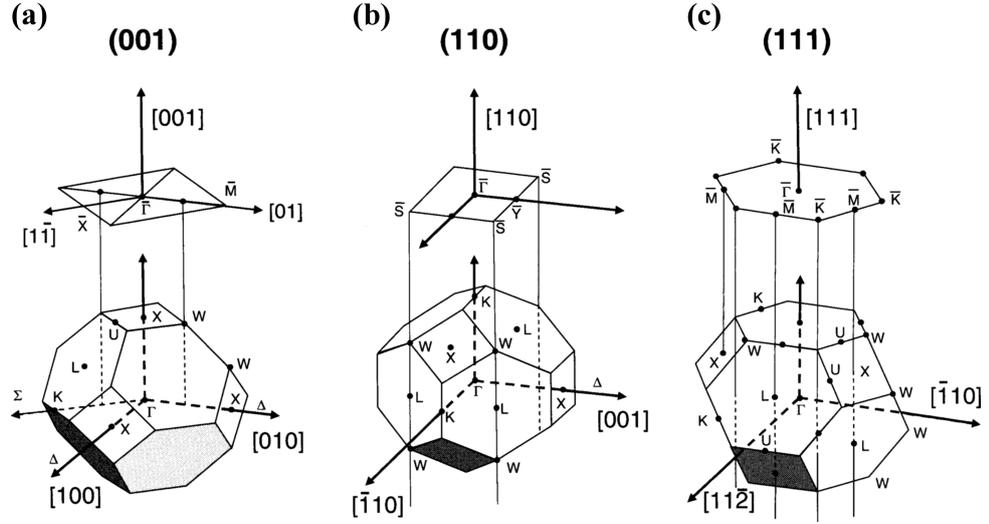


Figure 3.9. Schematic picture for the construction of the surface Brillouin zones by projecting the bulk Brillouin zone onto the surfaces with normal vectors in the [001], [110], and [111] directions, respectively. Figure taken from Ref. 90.

3.4.1.1 Intra-layer hopping

Similar to the tight-binding approach for the bulk crystal, we can write down a 20×20 intra-layer Bloch Hamiltonian $\mathcal{H}_{0\parallel}$ which describes the hopping of electrons within a single layer of HgTe or CdTe. It turns out that $\mathcal{H}_{0\parallel}$ is also block-diagonal in momentum space:

$$\mathcal{H}_{0\parallel} = \begin{bmatrix} \mathfrak{h}_{aa}^{0\parallel}(\mathbf{k}_{\parallel}) & 0 & \mathfrak{h}_{ac}^{0\parallel}(\mathbf{k}_{\parallel}) & 0 \\ 0 & \mathfrak{h}_{aa}^{0\parallel}(\mathbf{k}_{\parallel}) & 0 & \mathfrak{h}_{ac}^{0\parallel}(\mathbf{k}_{\parallel}) \\ \mathfrak{h}_{ca}^{0\parallel}(\mathbf{k}_{\parallel}) & 0 & \mathfrak{h}_{cc}^{0\parallel}(\mathbf{k}_{\parallel}) & 0 \\ 0 & \mathfrak{h}_{ca}^{0\parallel}(\mathbf{k}_{\parallel}) & 0 & \mathfrak{h}_{cc}^{0\parallel}(\mathbf{k}_{\parallel}) \end{bmatrix}. \quad (3.38)$$

Here, $\mathbf{k}_{\parallel} = (k_x, k_y, 0)^T$ denotes the conserved momentum within the **surface Brillouin zone** of the (001) plane, and the construction of such a surface Brillouin zone and its high-symmetry points is schematically shown in Fig. 3.9 for different orientations of the slab.

The diagonal block matrices $\mathfrak{h}_{\nu\nu}^{0\parallel}(\mathbf{k}_{\parallel})$ in $\mathcal{H}_{0\parallel}$ are independent of the momentum \mathbf{k}_{\parallel} , describing the relative energies of the orbital states with respect to each other:

$$\mathfrak{h}_{\nu\nu}^{0\parallel}(\mathbf{k}_{\parallel}) = \begin{bmatrix} E_{s\nu} & 0 & 0 & 0 & 0 \\ 0 & E_{p\nu} & 0 & 0 & 0 \\ 0 & 0 & E_{p\nu} & 0 & 0 \\ 0 & 0 & 0 & E_{p\nu} & 0 \\ 0 & 0 & 0 & 0 & E_{s^*\nu} \end{bmatrix}, \quad (3.39)$$

whereas the off-diagonal block matrices $\mathfrak{h}_{ac}^{0\parallel}(\mathbf{k}_{\parallel})$ describe the intra-layer interaction matrix elements between different orbitals on the anion and cation:

$$\mathfrak{h}_{ac}^{0\parallel}(\mathbf{k}_{\parallel}) = \begin{bmatrix} V_{ss} g_{0\parallel}(\mathbf{k}_{\parallel}) & V_{sp} g_{1\parallel}(\mathbf{k}_{\parallel}) & V_{sp} g_{2\parallel}(\mathbf{k}_{\parallel}) & V_{sp} g_{3\parallel}(\mathbf{k}_{\parallel}) & 0 \\ V_{ps} g_{1\parallel}(\mathbf{k}_{\parallel}) & V_{xx} g_{0\parallel}(\mathbf{k}_{\parallel}) & V_{xy} g_{3\parallel}(\mathbf{k}_{\parallel}) & V_{xy} g_{2\parallel}(\mathbf{k}_{\parallel}) & V_{ps^*} g_{1\parallel}(\mathbf{k}_{\parallel}) \\ V_{ps} g_{2\parallel}(\mathbf{k}_{\parallel}) & V_{xy} g_{3\parallel}(\mathbf{k}_{\parallel}) & V_{xx} g_{0\parallel}(\mathbf{k}_{\parallel}) & V_{xy} g_{1\parallel}(\mathbf{k}_{\parallel}) & V_{ps^*} g_{2\parallel}(\mathbf{k}_{\parallel}) \\ V_{ps} g_{3\parallel}(\mathbf{k}_{\parallel}) & V_{xy} g_{2\parallel}(\mathbf{k}_{\parallel}) & V_{xy} g_{1\parallel}(\mathbf{k}_{\parallel}) & V_{xx} g_{0\parallel}(\mathbf{k}_{\parallel}) & V_{ps^*} g_{3\parallel}(\mathbf{k}_{\parallel}) \\ 0 & V_{s^*p} g_{1\parallel}(\mathbf{k}_{\parallel}) & V_{s^*p} g_{2\parallel}(\mathbf{k}_{\parallel}) & V_{s^*p} g_{3\parallel}(\mathbf{k}_{\parallel}) & 0 \end{bmatrix}. \quad (3.40)$$

oriented slab geometry take the following form:

$$\mathfrak{h}_{ac}^{0\perp}(\mathbf{k}_{\parallel}) = \begin{bmatrix} V_{ss} g_{0\perp}(\mathbf{k}_{\parallel}) & V_{sp} g_{1\perp}(\mathbf{k}_{\parallel}) & V_{sp} g_{2\perp}(\mathbf{k}_{\parallel}) & V_{sp} g_{3\perp}(\mathbf{k}_{\parallel}) & 0 \\ V_{ps} g_{1\perp}(\mathbf{k}_{\parallel}) & V_{xx} g_{0\perp}(\mathbf{k}_{\parallel}) & V_{xy} g_{3\perp}(\mathbf{k}_{\parallel}) & V_{xy} g_{2\perp}(\mathbf{k}_{\parallel}) & V_{ps^*} g_{1\perp}(\mathbf{k}_{\parallel}) \\ V_{ps} g_{2\perp}(\mathbf{k}_{\parallel}) & V_{xy} g_{3\perp}(\mathbf{k}_{\parallel}) & V_{xx} g_{0\perp}(\mathbf{k}_{\parallel}) & V_{xy} g_{1\perp}(\mathbf{k}_{\parallel}) & V_{ps^*} g_{2\perp}(\mathbf{k}_{\parallel}) \\ V_{ps} g_{3\perp}(\mathbf{k}_{\parallel}) & V_{xy} g_{2\perp}(\mathbf{k}_{\parallel}) & V_{xy} g_{1\perp}(\mathbf{k}_{\parallel}) & V_{xx} g_{0\perp}(\mathbf{k}_{\parallel}) & V_{ps^*} g_{3\perp}(\mathbf{k}_{\parallel}) \\ 0 & V_{s^*p} g_{1\perp}(\mathbf{k}_{\parallel}) & V_{s^*p} g_{2\perp}(\mathbf{k}_{\parallel}) & V_{s^*p} g_{3\perp}(\mathbf{k}_{\parallel}) & 0 \end{bmatrix}, \quad (3.44)$$

where we have introduced another set of functions $g_{j\perp}(\mathbf{k}_{\parallel})$ with $j = 0, \dots, 3$. These functions are defined by a summation over those phase factors $\exp(i\mathbf{k}_{\parallel} \cdot \mathbf{d}_j)$ connecting the adjacent layers. Like the intra-layer functions $g_{j\parallel}(\mathbf{k}_{\parallel})$ the particular form of these functions depends on the orientation of the slab, and thus $g_{j\perp}(\mathbf{k}_{\parallel})$ has to take the symmetries of the slab into account. For example, for slabs with (001) surfaces, those function are given by

$$g_{0\perp}(\mathbf{k}_{\parallel}) \equiv \frac{1}{4}(e^{i\mathbf{k}_{\parallel} \cdot \mathbf{d}_3} + e^{i\mathbf{k}_{\parallel} \cdot \mathbf{d}_4}), \quad (3.45a)$$

$$g_{1\perp}(\mathbf{k}_{\parallel}) \equiv \frac{1}{4}(-e^{i\mathbf{k}_{\parallel} \cdot \mathbf{d}_3} - e^{i\mathbf{k}_{\parallel} \cdot \mathbf{d}_4}), \quad (3.45b)$$

$$g_{2\perp}(\mathbf{k}_{\parallel}) \equiv \frac{1}{4}(e^{i\mathbf{k}_{\parallel} \cdot \mathbf{d}_3} - e^{i\mathbf{k}_{\parallel} \cdot \mathbf{d}_4}), \quad (3.45c)$$

$$g_{3\perp}(\mathbf{k}_{\parallel}) \equiv \frac{1}{4}(-e^{i\mathbf{k}_{\parallel} \cdot \mathbf{d}_3} + e^{i\mathbf{k}_{\parallel} \cdot \mathbf{d}_4}). \quad (3.45d)$$

3.4.1.3 Spin-orbit interaction

Finally, the spin-orbit interaction coupling spin-up and spin-down states of different p orbitals on each atom is given by the same interaction matrix as was introduced for the bulk electronic band structure [cf. Eqs (3.25–3.27)]:

$$\mathcal{H}_{\text{SO}} = \begin{bmatrix} \mathfrak{h}_{aa}^{\text{SO}}(\uparrow\uparrow) & \mathfrak{h}_{aa}^{\text{SO}}(\uparrow\downarrow) & 0 & 0 \\ \mathfrak{h}_{aa}^{\text{SO}}(\downarrow\uparrow) & \mathfrak{h}_{aa}^{\text{SO}}(\downarrow\downarrow) & 0 & 0 \\ 0 & 0 & \mathfrak{h}_{cc}^{\text{SO}}(\uparrow\uparrow) & \mathfrak{h}_{cc}^{\text{SO}}(\uparrow\downarrow) \\ 0 & 0 & \mathfrak{h}_{cc}^{\text{SO}}(\downarrow\uparrow) & \mathfrak{h}_{cc}^{\text{SO}}(\downarrow\downarrow) \end{bmatrix}. \quad (3.46)$$

Since this is a local interaction, we take spin-orbit interactions in a finite-size slab geometry into account by adding the 20×20 matrix \mathcal{H}_{SO} to the diagonal elements of the Bloch Hamiltonian for the slab, $\mathcal{H}_{0\parallel}$.

3.4.2 Quantum wells with (001) oriented faces

Finally, to construct the full Bloch Hamiltonian matrix for a CdTe/HgTe/CdTe quantum well heterostructure consisting of N_{CdTe} layers of CdTe surrounding N_{HgTe} layers of HgTe on both sides, we proceed in a straightforward manner by combining the three super-matrices for the separate CdTe and HgTe (001) epilayers. Note that one has to properly connect these epilayers at the interfaces between CdTe and HgTe, depending on the termination of each of these layers. The interface between cadmium, tellurium, and mercury can be modeled by either an abrupt change of the empirical tight-binding parameters from one layer to next, or by a linear interpolation of the tight-binding parameters for tellurium. However, the electronic band structure of the quantum well structure turns out to be insensitive to the qualitative model details of the interface, but is rather determined by the bulk properties of the materials, in particular by the

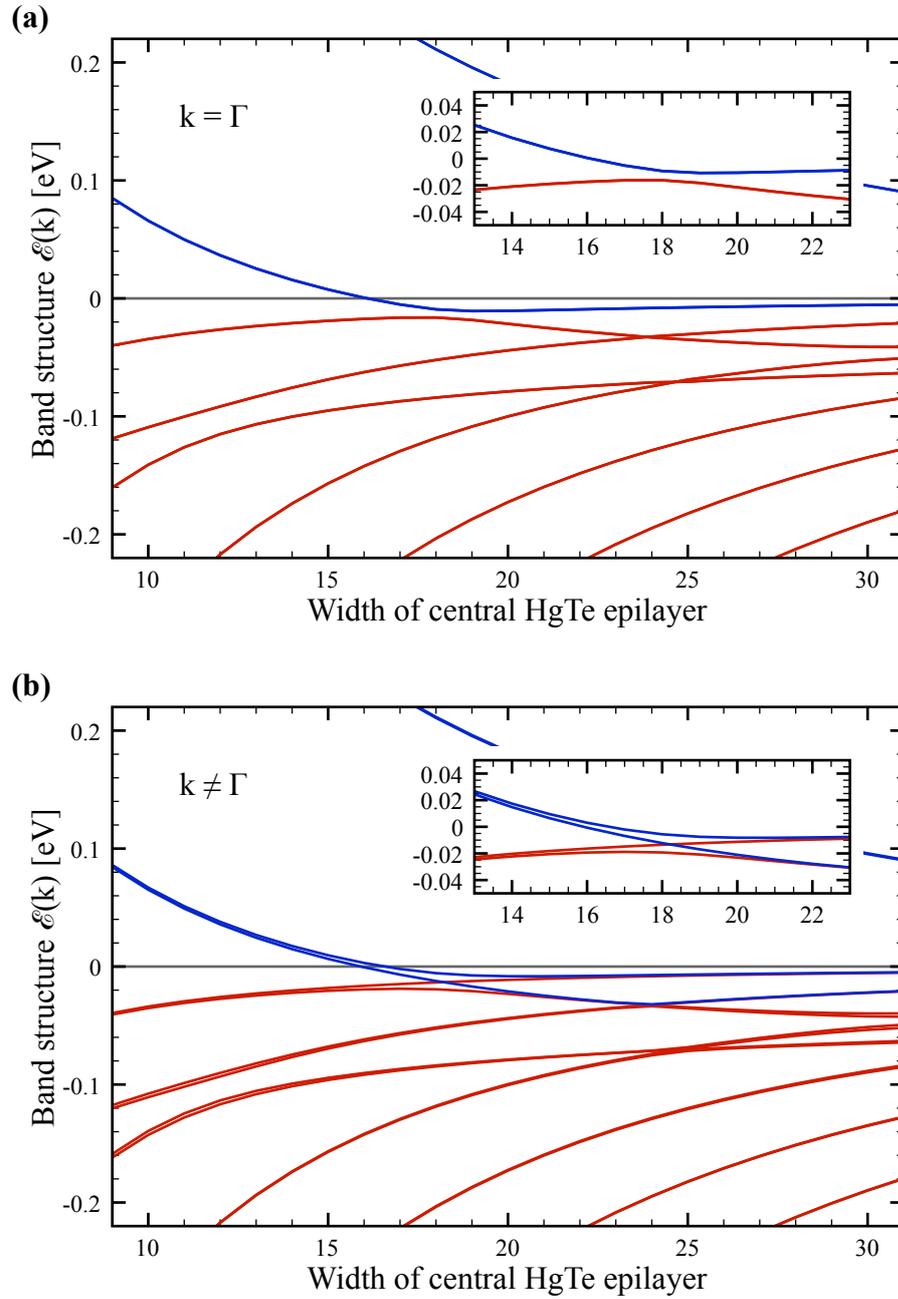


Figure 3.10. (a) Plot of the subband spectrum at the center of the Brillouin zone, *i.e.*, the Γ point, as function of the layer thickness N_{HgTe} for a (001) oriented CdTe/HgTe/CdTe quantum well with $N_{\text{CdTe}} = 8$ epilayers of CdTe surrounding HgTe on both sides. (b) Plot of the subband spectrum at one of the non-zero crossing points, where a crossing of the conduction and valence bands is observed. Note that away from the Γ point the double degeneracy of the subbands is removed due to the absence of inversion symmetry in (001) quantum wells. Depending on the overlap parameters and the valence band offset E_{VBO} we find certain points in the surface Brillouin zone, where the subbands cross as function of the quantum well thickness d . The exact position in the Brillouin zone, however, depends on quantitative details. Figure similar to Ref. 9.

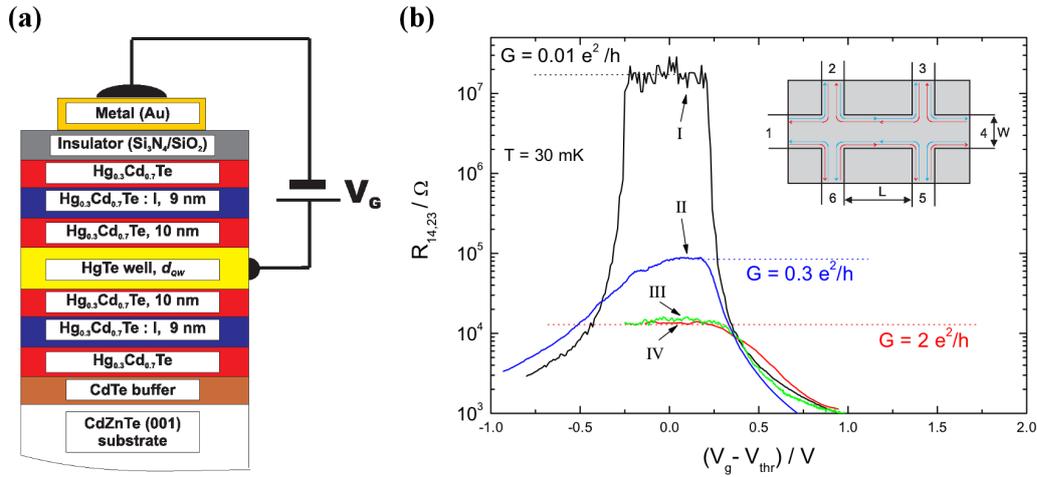


Figure 3.11. (a) Schematic illustration of the layered CdTe/HgTe/CdTe quantum well heterostructures grown epitaxially by MBE. (b) Longitudinal four-terminal resistance $R_{14,23}$ for different quantum well heterostructures as function of the gate voltage measured at $T = 30$ mK. Sample I with quantum well width $d = 5.5$ nm $< d_c$ is in the normal regime, while samples II, III, and IV with quantum well width $d = 7.3$ nm $> d_c$ show an inverted band structure. Most importantly, in the normal regime, the system resembles an ordinary insulator with a very large resistance. In the inverted regime, however, there exist topologically protected edge states running along the sample which lead to a quantized resistance of $2e^2/h$. The inset shows an illustration of the devices used by König *et al.*, where the edge states at the boundaries of the gated region (indicated by the red and blue lines) run in opposite direction. Figure after Refs. 6,27.

Finally, note that the situation in [001] quantum well heterostructure is somewhat special. For instance, a quantum well heterostructure grown on a substrate with (111) faces has the point group C_{3v} , and hence the surface states must always cross at the Γ point as a function of layer thickness, *i.e.*, one cannot observe the above anti-crossing of levels close to the Fermi level. To conclude, the character of the quantum well subbands and the corresponding band structure topology depends sensitively on the growth direction of the quantum well and its crystal symmetries⁹¹.

3.4.3 Experimental discovery of the quantum spin Hall insulator

In this section, let us briefly review some of the experimental observations of the transport properties of CdTe/HgTe/CdTe quantum wells made by König *et al.* (Refs. 6,27). Experimentally, they have investigated the properties of HgTe samples which have a quantum well width in the range 4.5 – 12 nm, and thereby they have studied samples both in the normal and the inverted regime. The samples grown by molecular beam epitaxy and structured by means of optical and electron lithography have a particular sequence of different layers made of CdTe and HgTe, as shown in Fig. 3.11 (a). Here, the HgTe quantum wells are surrounded by mercury-doped Hg_{0.3}Cd_{0.7}Te barriers and by iridium-doped layers, which allows to control the electron mobility in the samples. For example, König *et al.* report electron mobilities of several 10^{15} cm²/(V s) even at low densities $n < 5 \cdot 10^{11}$ cm⁻², while the elastic mean free path is of the order of several μ m, which opens up the possibility for transport measurements on those devices. For the investigation of the transport properties König *et al.* have used devices in a conventional six-terminal Hall bar geometry, with quantum well widths of about 4.5 – 12 nm.

In particular, the appearance of edge channels in the inverted quantum wells with thick-

ness $d > d_c = 6.3$ nm was studied in small six-terminal Hall bar geometries. The data shown in Fig. 3.11 (b) from Ref. 6 shows the four-terminal resistance $R_{14,23} = V_{23}/I_{14}$ as function of the normalized gate voltage for several devices. Apparently, the data for sample I with a quantum well width $d = 5.5$ nm $< d_c$ shows the behavior expected for a device with a normal band structure, because the resistance $R_{14,23}$ increases by many orders of magnitude when the Fermi level is in the gap. On the other hand, samples II, III, and IV show that the resistance remains finite in the insulating regime, *i.e.*, for $V_g - V_{\text{thr}} \approx 0$ V, which is the key signature of the edge states in the quantum spin Hall insulator phase. The inverse of the resistance, $G_{14,23}$, reaches a plateau close to the predicted value of $2e^2/h$ for the two edge states running in opposite directions. König *et al.* have also checked that the sample geometry does not affect this result by investigating samples of different dimensions [samples II, III, and IV in Fig. 3.11 (b)]. For further details on the fabrication of those CdTe/HgTe/CdTe quantum wells and experimental results see Refs. 5,6,27 and references therein.

3.5 Strained HgTe as three-dimensional topological insulator

The discovery of the two-dimensional quantum spin Hall insulator in CdTe/HgTe/CdTe quantum well heterostructures has generated a lot of interest in the topological properties of semiconductors, especially the generalization to three dimensions^{7,8,28,52,53,69–72}. However, many of those compounds under investigation, which have been predicted theoretically to realize 3D topological insulators, show strong defect doping and low carrier mobilities which makes the observation of charge transport on the surface quite difficult¹⁰. HgTe samples, on the other hand, can be produced with high quality and high carrier mobilities, thereby opening up the possibility to perform transport measurements. Furthermore, HgTe is charge-neutral when the Fermi energy lies exactly at the band touching point of the Γ_{8v} valence band and the Γ_{8c} conduction band, and shows a band inversion of the Γ_6 and Γ_8 bands due to strong spin-orbit interactions, as discussed above. In principle, one can thus expect that three-dimensional HgTe also hosts Dirac-like states on its two-dimensional surfaces, but since HgTe is a semimetal, those surface states are always coupled to metallic bulk states and thus difficult to observe. Nevertheless, the application of strain can open up a band gap between the Γ_{8c} and Γ_{8v} bands, so that strained 3D HgTe is expected to be a three-dimensional strong topological insulator⁸. Recently, Brüne *et al.* have provided experimental evidence by quantum Hall measurements that strained 3D HgTe is indeed a strong topological insulator and hosts 2D Dirac-like surface states¹⁰.

In this section, we discuss our numerical tight-binding calculations in the presence of strain and show that indeed a direct band gap at the Γ point of the Brillouin zone is opened due to strain. To be specific, following Ref. 9 we consider a HgTe bulk sample grown epitaxially on a CdTe substrate with normal vector in the [100] direction, where strain is applied within the (100) plane normal to the growth direction. As mentioned before, the lattice constants of HgTe and CdTe are almost identical to each other, $a_{\text{HgTe}} = 6.46$ Å and $a_{\text{CdTe}} = 6.48$ Å, and the critical thickness for lattice relaxation of the HgTe epilayer is about 200 nm (Ref. 10). This implies that for thinner HgTe samples ($d_{\text{HgTe}} \ll 200$ nm) the epilayer adopts the in-plane lattice constant of the CdTe substrate which can be described in terms of **biaxial strain** within the (100) plain. Although the lattice mismatch between CdTe and HgTe is rather small,

$$\Delta a = \frac{a_{\text{CdTe}} - a_{\text{HgTe}}}{a_{\text{CdTe}}} \approx 0.3\%, \quad (3.47)$$

the resulting biaxial strain is sufficiently large to open a band gap between the Γ_{8v} and Γ_{8c}

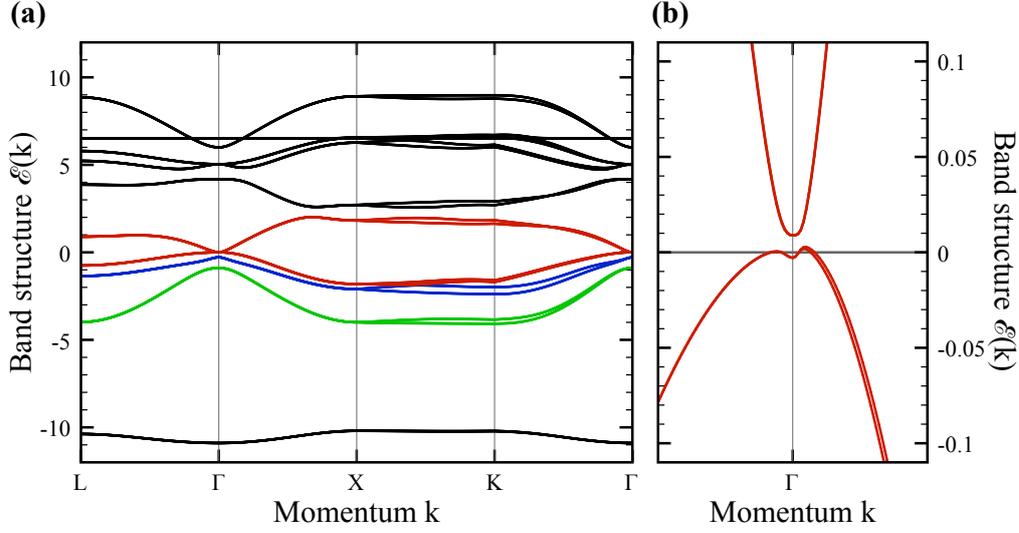


Figure 3.12. (a) Band structure $\mathcal{E}(\mathbf{k})$ of mercury telluride (HgTe) under biaxial strain generated by the lattice mismatch $\Delta a = 0.3\%$ as function of crystal momentum \mathbf{k} along the principal symmetry directions. The high-symmetry points are denoted by $L = \frac{2\pi}{a}(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})^T$, $\Gamma = (0, 0, 0)^T$, $X = \frac{2\pi}{a}(1, 0, 0)^T$, and $K = \frac{2\pi}{a}(\frac{3}{4}, \frac{3}{4}, 0)^T$, while the three high-symmetry directions $[100]$, $[110]$, and $[111]$ in the Brillouin zone are denoted by Δ , Σ and Λ . The so-called Γ_6 and Γ_8 valence and conduction bands are shown in red and blue color, respectively, while the spin-orbit split-off Γ_7 valence band is indicated by green. (b) Zoom into the band structure close to the Γ point. The applied strain leads to a direct energy gap of about $\mathcal{E}_{\text{gap}}^{\text{direct}}(\Gamma) = 12$ meV between the Γ_{8v} and Γ_{8c} bands.

bands. Evaluating the strain tensor for biaxial strain in the (100) plane (cf. appendix B for details) we have calculated the electronic band structure of strained HgTe. At first glance, the band structure shown in Fig. 3.12 (a) does not differ very much from the one without strain shown previously in Fig. 3.4. However, a closer inspection of the band structure at the Γ point [see Fig. 3.12 (b)] reveals a **direct band gap** instead of a band touching,

$$\mathcal{E}_{\text{gap}}^{\text{direct}}(\mathbf{k} = \Gamma) = 12 \text{ meV}, \quad (3.48)$$

which is in qualitative agreement with the $\mathbf{k} \cdot \mathbf{p}$ calculations by Brüne *et al.* (Ref. 10). Moreover, we also obtain a finite **indirect band gap** which is calculated from the bulk density of states of strained HgTe [see Fig. 3.13 (a)]:

$$\mathcal{E}_{\text{gap}}^{\text{indirect}} = 4 \text{ meV}. \quad (3.49)$$

Note that the indirect band gap is a factor three smaller than the direct band $\mathcal{E}_{\text{gap}}^{\text{direct}}(\Gamma)$. As a consequence, the fully strained 3D HgTe is insulating in the bulk, and thus the 2D surface Dirac states are no longer coupled to metallic bulk states. Finally, this opens up the possibility to study those Dirac-like surface states both experimentally and theoretically.

In the presence of uniaxial compressive strain along the $[001]$ direction, Dai *et al.* have previously shown that strained 3D HgTe is a strong topological insulator in three dimensions under the definitions given by Fu, Kane, and Mele^{3,7} as long as the bulk gap does not close⁹. Furthermore, they have explicitly calculated the topological surface states of strained 3D HgTe using the so-called **six-band Kane model** which is an effective low-energy Hamiltonian in terms of the Γ_{8v} and Γ_{7v} valence bands and the Γ_{6c} conduction band obtained by integrating out the effect of the other bands [see Fig. 3.14 (a)]. This approach is based on the $\mathbf{k} \cdot \mathbf{p}$ approximation

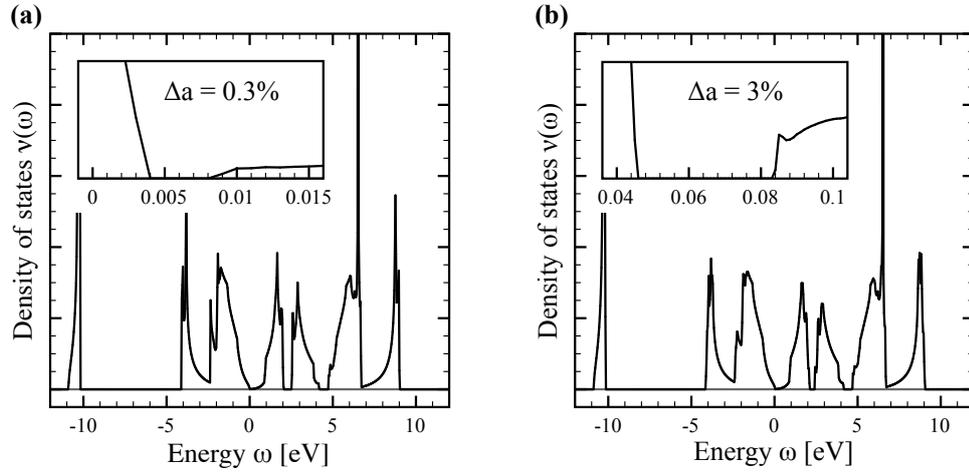


Figure 3.13. Density of states for strained HgTe with biaxial strain in the (100) plane generated by the lattice mismatch Δa , calculated numerically from the electronic band structure of bulk HgTe by performing exact diagonalization on $5 \cdot 10^9$ random sampling points in momentum space. In panel (a) $\Delta a = 0.3\%$ with the indirect band gap $\mathcal{E}_{\text{gap}}^{\text{indirect}} = 4$ meV, while in panel (b) the lattice mismatch is artificially increased to $\Delta a = 3\%$, leading to a ten times larger indirect band gap.

and thus only valid in the vicinity of the expansion point of the Taylor series which is usually taken as the center of the Brillouin zone, $\mathbf{k} = 0$.

It turns out that the origin of the topological surface states can be traced back to the band inversion of HgTe, in particular between the light-hole (LH) conduction band Γ_{8c} and the electron-like (E) valence band Γ_{8v} . The additional heavy-hole (HH) valence band is split from the LH band by an energy gap generated by the applied strain. However, the heavy-hole band is only important for details in the electronic band structure, but does not change the topology of the bands, as shown in Fig. 3.14 (b). Even when the heavy-hole band is fully coupled to and hybridizing with the E band, the strained 3D HgTe is a strong topological insulator with a single surface Dirac state crossing at the Γ point. In the next section, we construct a minimal model for 3D HgTe based on the electron (E) and light-hole (LH) states. Based on this model, we study in chapter 4 the quantum Hall effect of such a topological insulator in the presence of an external magnetic field¹⁹.

3.6 Construction of a minimal model for topological insulators

To develop an explicit model of a strong topological insulator, we consider a strained 3D HgTe sample and describe its low-energy properties in terms of the relevant states close to the Fermi level. In the following chapter, we then consider the quantum Hall effect of this model in the presence of an applied magnetic field, and discuss the relation between the quantized Hall conductance and the topological θ -term¹⁹.

3.6.1 Choice of basis states

As discussed in the previous section, HgTe exhibits a strong topological insulator phase under applied strain which opens up a direct band gap at the Γ point¹⁰. Dai *et al.* showed⁹ that the topological surface states of strained 3D HgTe can be described in terms of the Γ_6 conduction (E) bands and the Γ_8 light-hole (LH) bands while the Γ_8 heavy-hole (HH) bands can be ne-

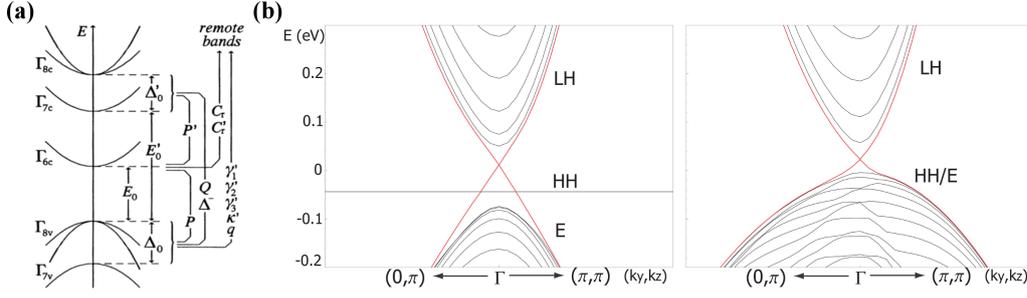


Figure 3.14. (a) Schematic band structure for the extended 14×14 Kane model with usual band ordering (Γ_{8v} band below Γ_{6c} band). For a definition of the $\mathbf{k} \cdot \mathbf{p}$ couplings see Ref. 94. (b) Band structure close to the Γ point calculated from the six-band Kane model for the decoupled heavy-hole (HH) band (left panel) and the fully coupled heavy-hole band (right panel). The two-dimensional surface states traversing the bulk gap (red lines) are clearly visible and show Dirac-like behavior at the Γ point, implying that only the electron (E) and the light-hole (LH) bands are important for the topological properties of strained 3D HgTe. Figure after Ref. 9.

glected. To construct an effective model for a strong topological insulator based on strained 3D HgTe, we limit ourselves to a four-dimensional basis set of wave functions in the following sequence, consisting only of those Γ_6 and Γ_8 states:

$$|1\rangle \equiv |\Gamma_6, +\frac{1}{2}\rangle = |E, +\frac{1}{2}\rangle = |s, \uparrow\rangle, \quad (3.50a)$$

$$|2\rangle \equiv |\Gamma_6, -\frac{1}{2}\rangle = |E, -\frac{1}{2}\rangle = |s, \downarrow\rangle, \quad (3.50b)$$

$$|3\rangle \equiv |\Gamma_8, +\frac{1}{2}\rangle = |LH, +\frac{1}{2}\rangle = \frac{1}{\sqrt{6}} (|p_x, \downarrow\rangle + i|p_y, \downarrow\rangle - 2|p_z, \uparrow\rangle), \quad (3.50c)$$

$$|4\rangle \equiv |\Gamma_8, -\frac{1}{2}\rangle = |LH, -\frac{1}{2}\rangle = \frac{-1}{\sqrt{6}} (|p_x, \uparrow\rangle - i|p_y, \uparrow\rangle + 2|p_z, \downarrow\rangle). \quad (3.50d)$$

The goal of this section is to derive the effective description of the relevant E and LH states for strained 3D HgTe and to construct a 4×4 Hamiltonian for a strong topological insulator using the theory of invariants which states that the Hamiltonian must be invariant under all symmetry operations of the system under consideration^{74,75}. To be specific, we consider a simple cubic lattice in three spatial dimensions, where the electronic states of the effective model are given by the set (3.50). In other words, we consider a system that is invariant with respect to spatial inversion (\hat{P}) and two-fold rotations about the x , y , and z axes ($\hat{\mathcal{R}}_x(\pi)$, $\hat{\mathcal{R}}_y(\pi)$, and $\hat{\mathcal{R}}_z(\pi)$). Furthermore, in the absence of time-reversal symmetry-breaking perturbations such as an external magnetic field, we consider a system that is invariant with respect to time-reversal ($\hat{\Theta}$). In the following, we briefly discuss the transformation of the basis set of wave functions (3.50) under those symmetries, and then construct a suitable Bloch Hamiltonian which respects the aforementioned symmetries (see Table 3.6). Finally, we discuss the bulk band structure and the phase diagram of that model.

3.6.2 Time-reversal symmetry

Let us first discuss the effect of time-reversal symmetry $\hat{\Theta}$ on the set of wave functions (3.50). We recall from quantum mechanics that the time-reversal operation for spin 1/2 fermions is represented by an anti-unitary operator $\hat{\Theta}$ and has a two-dimensional representation with the property $\hat{\Theta}^2 = -\mathbb{1}$, where $\mathbb{1}$ is the unit operator. For a spin 1/2 fermion, time-reversal is usually implemented by the representation $\hat{\Theta} = i\sigma_y \mathcal{K}$, where σ_y is the usual Pauli matrix, and

\mathcal{K} denotes complex conjugation. Thus, time-reversal flips the spin of the electron with a spin-dependent prefactor ± 1 for spin-up and spin-down states, respectively. In that sense, the spin-up and spin-down states are time-reversal partners or **Kramers' partners**.

It is a straightforward task to compute the transformation of the basis set of wave functions (3.50) under time reversal. As a result, the conduction band states $|E, \pm\frac{1}{2}\rangle$ states and the light-hole valence band states $|LH, \pm\frac{1}{2}\rangle$ are both Kramers' partners, *i.e.*,

$$\hat{\Theta}|E, \pm\frac{1}{2}\rangle = \pm|E, \mp\frac{1}{2}\rangle, \quad (3.51a)$$

$$\hat{\Theta}|LH, \pm\frac{1}{2}\rangle = \pm|LH, \mp\frac{1}{2}\rangle. \quad (3.51b)$$

Thus, in terms of the basis set (3.50) time-reversal may be represented by

$$\hat{\Theta} = i\tau_0 \otimes \sigma_y \mathcal{K} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{bmatrix} \mathcal{K}. \quad (3.52)$$

As above $\boldsymbol{\tau} = (\tau_x, \tau_y, \tau_z)^T$ is a vector of Pauli matrices acting on the orbital degrees of freedom, $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)^T$ acts on the spin degrees of freedom, and $\tau_0 = \sigma_0 = \mathbb{1}$ denotes the 2×2 identity matrix.

3.6.3 Spatial inversion symmetry

While time-reversal symmetry relates states with opposite spin to each other, the spatial inversion operation $\hat{\mathcal{P}}$ defines the parity of each state. First, recall that in Cartesian coordinates the inversion operation acts as $\hat{\mathcal{P}}(x, y, z) = (-x, -y, -z)$, whereas in spherical coordinates $\hat{\mathcal{P}}(r, \theta, \phi) = (r, \pi - \theta, \phi + \pi)$. As a consequence, the spherical harmonics $Y_{lm}(\theta, \phi)$ as the building blocks of the s and p orbital wave functions transform as:

$$\hat{\mathcal{P}}Y_{lm}(\theta, \phi) = Y_{lm}(\pi - \theta, \phi + \pi) = (-1)^l Y_{lm}(\theta, \phi). \quad (3.53)$$

This implies that the conduction band states $|E, \pm\frac{1}{2}\rangle$ have even parity, because they stem from s orbital ($l = 0$) states. On the other hand, the light-hole valence band states $|LH, \pm\frac{1}{2}\rangle$ are superpositions of p orbital states, and thus they are odd under spatial inversion. Hence, in terms of the basis set (3.50) $\hat{\mathcal{P}}$ may be represented by a diagonal 4×4 matrix:

$$\hat{\mathcal{P}} = \tau_z \otimes \sigma_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \quad (3.54)$$

with eigenvalues ± 1 . Note that the parity of the different states can greatly simplify the calculation of the effective 4×4 Hamiltonian in the following.

3.6.4 Rotations about the Cartesian axes

Besides the usual temporal and spatial inversion symmetries, we also consider certain rotational symmetries of the basis states (3.50), in particular the symmetries of the wave functions with respect to rotations about the Cartesian axes.

3.6.4.1 Two-fold rotations about the x axis

Let us first discuss two-fold rotations about the x axis, denoted by $\hat{\mathcal{R}}_x(\pi)$. In Cartesian coordinates, the rotation by π within the yz plane reverses the sign of the y and z coordinates, $\hat{\mathcal{R}}_x(\pi)(x, y, z) = (x, -y, -z)$, while in spherical coordinates we have $\hat{\mathcal{R}}_x(\pi)(r, \theta, \phi) = (r, \pi - \theta, -\phi)$. Hence, the spherical harmonics $Y_{lm}(\theta, \phi)$ transform under two-fold rotations about the x axis as

$$\hat{\mathcal{R}}_x(\pi)Y_{lm}(\theta, \phi) = Y_{lm}(\pi - \theta, -\phi) = (-1)^l Y_{l, -m}(\theta, \phi). \quad (3.55)$$

Note that in contrast to the spatial inversion operation the rotation about the x axis interchanges the corresponding spherical harmonics Y_{lm} and $Y_{l, -m}$.

However, aside from the orbital part of the wave function we also have to consider the spin degree of freedom. From quantum mechanics we recall that for a spin $1/2$ electron a rotation about an arbitrary axis $\hat{\mathbf{n}}$ by the angle α is described in terms of a unitary operator $\mathcal{U}(\hat{\mathbf{n}}, \alpha)$ defined by

$$\mathcal{U}(\hat{\mathbf{n}}, \alpha) \equiv \exp\left(\frac{i\alpha}{2\hbar} \hat{\mathbf{n}} \cdot \boldsymbol{\sigma}\right) = \sigma_0 \cos(\alpha/2) + i(\hat{\mathbf{n}} \cdot \boldsymbol{\sigma}) \sin(\alpha/2). \quad (3.56)$$

As a consequence, the two pairs of conduction band states $\{|E, \frac{1}{2}\rangle, |E, -\frac{1}{2}\rangle\}$ and light-hole valence band states $\{|LH, \frac{1}{2}\rangle, |LH, -\frac{1}{2}\rangle\}$ are partners under the rotation operation $\hat{\mathcal{R}}_x(\pi)$:

$$\hat{\mathcal{R}}_x(\pi)|\sigma\rangle = e^{i\pi/2}|\bar{\sigma}\rangle \quad (3.57)$$

with $\sigma \in \{\uparrow, \downarrow\}$ and $\bar{\sigma} \in \{\downarrow, \uparrow\}$, respectively. Combining the transformation properties of the orbital part and the spin part of the wave function, the rotation operation $\hat{\mathcal{R}}_x(\pi)$ acting on the basis set (3.50) yields

$$\hat{\mathcal{R}}_x(\pi)|E, \pm\frac{1}{2}\rangle = e^{i\pi/2}|E, \mp\frac{1}{2}\rangle, \quad (3.58a)$$

$$\hat{\mathcal{R}}_x(\pi)|LH, \pm\frac{1}{2}\rangle = e^{i\pi/2}|LH, \mp\frac{1}{2}\rangle. \quad (3.58b)$$

In terms of a skew-Hermitian (or anti-Hermitian) 4×4 matrix, $\hat{\mathcal{R}}_x(\pi)$ takes the form

$$\hat{\mathcal{R}}_x(\pi) = i\tau_z \otimes \sigma_x = \begin{bmatrix} 0 & i & 0 & 0 \\ i & 0 & 0 & 0 \\ 0 & 0 & 0 & -i \\ 0 & 0 & -i & 0 \end{bmatrix}. \quad (3.59)$$

Note that the two-fold application of $\hat{\mathcal{R}}_x(\pi)$, *i.e.*, a full revolution around the x axis, leads to a sign change of the wave function. This can be understood from the transformation behavior of the electronic spin which rotated by 2π leads to the phase factor $e^{i\pi} = -1$, leading to the well-known sign change of the spin wave function when the electron spin performs a full revolution.

3.6.4.2 Two-fold rotations about the y axis

We also consider two-fold rotations about the y axis, denoted by $\hat{\mathcal{R}}_y(\pi)$. Obviously, the action of $\hat{\mathcal{R}}_y(\pi)$ in Cartesian coordinates is given by $\hat{\mathcal{R}}_y(\pi)(x, y, z) = (-x, y, -z)$, while in spherical coordinates $\hat{\mathcal{R}}_y(\pi)(r, \theta, \phi) = (r, \pi - \theta, \pi - \phi)$. Thus, the spherical harmonics $Y_{lm}(\theta, \phi)$ transform under two-fold rotations about the y axis as

$$\hat{\mathcal{R}}_y(\pi)Y_{lm}(\theta, \phi) = Y_{lm}(\pi - \theta, \pi - \phi) = (-1)^{l+m} Y_{l, -m}(\theta, \phi). \quad (3.60)$$

Just as for rotations about the x axis, the spherical harmonics Y_{lm} and $Y_{l,-m}$ are partners of each other, but in contrast to $\hat{\mathcal{R}}_x(\pi)$ there exists another phase factor $(-1)^m = e^{im\pi}$ which stems from the transformation of the azimuthal angle ϕ under $\hat{\mathcal{R}}_y(\pi)$. Taking into account the rotation of the electron spin about the y axis as described by $\mathcal{U}(\hat{\mathbf{n}}, \pi)$ with $\hat{\mathbf{n}} = \mathbf{e}_y$, the basis set (3.50) transforms under the rotation operation $\hat{\mathcal{R}}_y(\pi)$ as

$$\hat{\mathcal{R}}_y(\pi)|E, \pm\frac{1}{2}\rangle = \pm|E, \mp\frac{1}{2}\rangle, \quad (3.61a)$$

$$\hat{\mathcal{R}}_y(\pi)|LH, \pm\frac{1}{2}\rangle = \pm|LH, \mp\frac{1}{2}\rangle. \quad (3.61b)$$

This leads to the following skew-symmetric 4×4 matrix:

$$\hat{\mathcal{R}}_y(\pi) = i \tau_z \otimes \sigma_y = \begin{bmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \quad (3.62)$$

In analogy to $\hat{\mathcal{R}}_x(\pi)$, the two-fold application of $\hat{\mathcal{R}}_y(\pi)$ results in a sign change of the wave function due to the electron spin acquiring the phase factor $e^{i\pi} = -1$.

3.6.4.3 Two-fold rotations about the z axis

Finally, let us discuss two-fold rotations about the z axis, $\mathcal{R}_z(\pi)$. In spherical coordinates, $\hat{\mathcal{R}}_z(\pi)(r, \theta, \phi) = (r, \theta, \phi + \pi)$, so that the spherical harmonic $Y_{lm}(\theta, \phi)$ change sign:

$$\hat{\mathcal{R}}_z(\frac{\pi}{2})Y_{lm}(\theta, \phi) = Y_{lm}(\theta, \phi + \pi) = (-1)^m Y_{lm}(\theta, \phi). \quad (3.63)$$

Thus, in contrast to rotations about the x or y axes, the spherical harmonics Y_{lm} are eigenfunctions of the rotation operation $\mathcal{R}_z(\pi)$ with eigenvalues $(-1)^m$. Moreover, since the quantization axis of the electron spin coincides with the rotation axis, the spin part of the wave function is diagonal as well with eigenvalues $e^{\pm i\pi/2} = \pm i$ for spin-up and spin-down states, respectively. As a consequence, the transformation of the basis set (3.50) under two-fold rotations about the z axis is given by:

$$\hat{\mathcal{R}}_z(\pi)|E, \pm\frac{1}{2}\rangle = \pm i|E, \pm\frac{1}{2}\rangle, \quad (3.64a)$$

$$\hat{\mathcal{R}}_z(\pi)|LH, \pm\frac{1}{2}\rangle = \pm i|LH, \pm\frac{1}{2}\rangle. \quad (3.64b)$$

Therefore, the rotation operation $\mathcal{R}_z(\pi)$ can be represented as a complex diagonal 4×4 matrix:

$$\hat{\mathcal{R}}_z(\pi) = i \tau_0 \otimes \sigma_z = \begin{bmatrix} i & 0 & 0 & 0 \\ 0 & -i & 0 & 0 \\ 0 & 0 & i & 0 \\ 0 & 0 & 0 & -i \end{bmatrix}. \quad (3.65)$$

This completes the discussion of symmetry operations for the minimal model.

3.6.5 Effective Hamiltonian for a topological insulator based on symmetries

After the brief discussion of the symmetry operations, we now construct a minimal Hamiltonian for topological insulators in three dimensions based on the four-state basis of HgTe describing conduction band states (denoted by E) and light-hole valence band states (denoted

by LH) [cf. Eq. (3.50)]. As discussed in the previous subsections, we consider a system that has time-reversal symmetry ($\hat{\Theta}$), spatial inversion symmetry ($\hat{\mathcal{P}}$), and two-fold rotational symmetries about the Cartesian coordinate axes ($\hat{\mathcal{R}}_x(\pi)$, $\hat{\mathcal{R}}_y(\pi)$, and $\hat{\mathcal{R}}_z(\pi)$).

In general, any 4×4 Hamiltonian matrix can be expressed in terms of the identity matrix $\mathbb{1}$, five Γ -matrices² Γ_a , and ten commutators Γ_{ab} thereof as

$$\mathcal{H}(\mathbf{k}) = \epsilon(\mathbf{k})\mathbb{1} + \sum_{a=0}^4 d_a(\mathbf{k})\Gamma_a + \sum_{a<b=0}^4 d_{ab}(\mathbf{k})\Gamma_{ab}, \quad (3.66)$$

where the five 4×4 matrices $\Gamma_{0,\dots,4}$ satisfy the usual Clifford algebra:

$$\{\Gamma_a, \Gamma_b\} = 2\delta_{ab}\mathbb{1} \quad \text{with} \quad a, b = 0, \dots, 4, \quad (3.67)$$

and δ_{ab} denotes the Kronecker delta. Moreover, the ten non-vanishing Γ -matrices Γ_{ab} are defined by commutators of Γ_a and Γ_b :

$$\Gamma_{ab} \equiv \frac{1}{2i} [\Gamma_a, \Gamma_b]. \quad (3.68)$$

Since the Bloch Hamiltonian matrix should be invariant under the set of symmetry operations $\{\hat{\Theta}, \hat{\mathcal{P}}, \hat{\mathcal{R}}_x(\pi), \hat{\mathcal{R}}_y(\pi), \hat{\mathcal{R}}_z(\pi)\}$, the functions $d_a(\mathbf{k})$ and $d_{ab}(\mathbf{k})$ should have the same transformation behavior under those symmetry operations as Γ_a and Γ_{ab} , respectively. Therefore, we have to work out the transformation form of the Γ matrices under the above symmetry operations. To be specific, we construct the Γ matrices as follows:

$$\{\Gamma_0, \Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4\} \equiv \{\tau_z \otimes \sigma_0, -\tau_x \otimes \sigma_x, -\tau_x \otimes \sigma_y, -\tau_x \otimes \sigma_z, \tau_y \otimes \sigma_0\}. \quad (3.69)$$

As defined above, τ_α and σ_α with $\alpha \in \{x, y, z\}$ denote Pauli matrices acting on the orbital and spin degrees of freedom, respectively. Since $\tau_\alpha^2 = \sigma_\alpha^2 = \mathbb{1}$ for all $\alpha \in \{x, y, z\}$ it is straightforward to show that

$$\Gamma_a^2 = \Gamma_{ab}^2 = \mathbb{1}, \quad \{\Gamma_a, \Gamma_{ab}\} = 0, \quad \{\Gamma_{ab}, \Gamma_{ac}\} = 0 \quad \text{for} \quad b \neq c. \quad (3.70)$$

For the above Γ matrices, we have calculated the transformation behavior under the symmetry transformations $\hat{\Theta}$, $\hat{\mathcal{P}}$, $\hat{\mathcal{R}}_x(\pi)$, $\hat{\mathcal{R}}_y(\pi)$, and $\hat{\mathcal{R}}_z(\pi)$ which are listed in Table 3.6 for reference. For example, a term like $\sin(k_x a)$ is odd under spatial and temporal inversion as well as under two-fold rotations about the y and z axes, but even under two-fold rotations about the x axis itself. A close inspection of Table 3.6 shows that only Γ_1 has the same signature under all those symmetry operations, and thus the Bloch Hamiltonian contains a term $\sin(k_x a)\Gamma_1$. For a system that preserves the symmetries generated by $\hat{\Theta}$, $\hat{\mathcal{P}}$, $\hat{\mathcal{R}}_x(\pi)$, $\hat{\mathcal{R}}_y(\pi)$, and $\hat{\mathcal{R}}_z(\pi)$ we therefore obtain the following general Bloch Hamiltonian $\mathcal{H}(\mathbf{k})$:

$$\mathcal{H}_0(\mathbf{k}) = \epsilon(\mathbf{k})\mathbb{1} + m(\mathbf{k})\Gamma_0 - t \sum_{j=1}^3 (\cos(k_j a)\Gamma_0 + \sin(k_j a)\Gamma_j). \quad (3.71)$$

Here, $\epsilon(\mathbf{k})$ describes the global bending of the bands throughout the Brillouin zone, $m(\mathbf{k})$ denotes the **mass parameter** or **tuning parameter** which allows to realize different topological phases of the model, and t is the **overlap parameter** or **hopping amplitude**. In the following, we neglect the global bending of the bands, *i.e.*, we set $\epsilon(\mathbf{k}) = 0$, and we consider only a constant tuning parameter, *i.e.*, we consider $m(\mathbf{k}) \equiv m = \text{const.}$

²Unfortunately, the time-reversal invariant momenta introduced by Kane and Mele are also labeled by the symbol “ Γ ” with subscript “ i ” which should not be confused with the Γ_a matrices introduced for the minimal model or the center of the Brillouin zone $\mathbf{k} = \Gamma$.

	$\hat{\Theta}$	$\hat{\mathcal{P}}$	$\hat{\mathcal{R}}_x(\pi)$	$\hat{\mathcal{R}}_y(\pi)$	$\hat{\mathcal{R}}_z(\pi)$
Γ_0	+	+	+	+	+
Γ_1	-	-	+	-	-
Γ_2	-	-	-	+	-
Γ_3	-	-	-	-	+
Γ_4	-	-	+	+	+
Γ_{01}	+	-	+	-	-
Γ_{02}	+	-	-	+	-
Γ_{03}	+	-	-	-	+
Γ_{04}	+	-	+	+	+
Γ_{12}	-	+	-	-	+
Γ_{13}	-	+	-	+	-
Γ_{14}	-	+	+	-	-
Γ_{23}	-	+	+	-	-
Γ_{24}	-	+	-	+	-
Γ_{34}	-	+	-	-	+

Table 3.6. List of the signatures of the Γ matrices [Eq. (3.69)] under time reversal $\hat{\Theta}$ [cf. Eq. (3.52)], spatial inversion $\hat{\mathcal{P}}$ [cf. Eq. (3.54)], and rotations about the Cartesian axes $\hat{\mathcal{R}}_x(\pi)$, $\hat{\mathcal{R}}_y(\pi)$, and $\hat{\mathcal{R}}_z(\pi)$ [cf. Eqs. (3.59, 3.62, 3.65)]. The “ \pm ” symbol indicates that the matrix has positive (negative) signature defined by, for example, $\hat{\Theta}\Gamma_0\hat{\Theta}^{-1} = +\Gamma_0$.

Additional terms which preserve certain rotation symmetries, but break time-reversal symmetry and/or inversion symmetry can be easily identified from Table 3.6 by close inspection. The first term, $\Delta\Gamma_4$, breaks both time-reversal and inversion symmetry. For example, the application of an external magnetic field breaks time-reversal symmetry, and thus it induces a small, but finite contribution to $\Delta\Gamma_4$. In contrast, the term $\Delta'\Gamma_{04}$ only breaks inversion symmetry, and such a term will be present in any system without an inversion center such as HgTe. Concerning the investigation of the quantum Hall effect in the next chapter we will usually focus on the term $\Delta\Gamma_{04}$ and neglect $\Delta'\Gamma_{04}$ in the following, but for the sake of completeness we briefly discuss the effect of $\Delta'\Gamma_{04}$ on the band structure as well.

To conclude, in this thesis we mainly study the following tight-binding Hamiltonian:

$$\mathcal{H}(\mathbf{k}) = m\Gamma_0 - t \sum_{j=1}^3 (\cos(k_j a)\Gamma_0 + \sin(k_j a)\Gamma_j) + \Delta\Gamma_4. \quad (3.72)$$

By inverse Fourier decomposition of the four-component spinors $\Psi_{\mathbf{k}}$,

$$\Psi_{\mathbf{R}} = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{R}} \Psi_{\mathbf{k}}, \quad \Psi_{\mathbf{k}} = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}} e^{-i\mathbf{k}\cdot\mathbf{R}} \Psi_{\mathbf{R}}, \quad (3.73)$$

we can easily calculate the real-space Hamiltonian of the 3D topological insulator on a square lattice which can be formulated as

$$H = m \sum_{\mathbf{R}} \Psi_{\mathbf{R}}^\dagger \Gamma_0 \Psi_{\mathbf{R}} - t \sum_{\mathbf{R}} \sum_{j=1}^3 \left[\Psi_{\mathbf{R}}^\dagger \left(\frac{\Gamma_0 - i\Gamma_j}{2} \right) \Psi_{\mathbf{R}-\mathbf{e}_j} + \text{H.c.} \right] + \Delta \sum_{\mathbf{R}} \Psi_{\mathbf{R}}^\dagger \Gamma_4 \Psi_{\mathbf{R}}. \quad (3.74)$$

This lattice Dirac Hamiltonian in 3+1 dimensions can be generalized to a time-reversal invariant topological insulator in 4+1 dimensions by adding a hopping term $\Psi_{\mathbf{R}}^\dagger (\Gamma_0 - i\Gamma_4) \Psi_{\mathbf{R}} + \text{H.c.}$ to

the Hamiltonian (3.74)¹¹. This generalization, however, works only up to four dimensions due to the fact that a four-band Hamiltonian can be constructed from just five Γ matrices $\Gamma_{0,\dots,4}$ and commutators thereof.

Finally, note that we can also write the Bloch Hamiltonian (3.72) in the following form:

$$\mathcal{H}(\mathbf{k}) = m \Gamma_0 - t \sum_{j=1}^3 (\cos(k_j a) \Gamma_0 + \sin(k_j a) \Gamma_j) + \Delta \Gamma_4 = \mathbf{d}(\mathbf{k}) \cdot \mathbf{\Gamma}, \quad (3.75)$$

where we have introduced the five-component vectors $\mathbf{d}(\mathbf{k})$ and $\mathbf{\Gamma}$ given by

$$\mathbf{d}(\mathbf{k}) \equiv (m - t \cos(k_x a) - t \cos(k_y a) - t \cos(k_z a), \\ - t \sin(k_x a), -t \sin(k_y a), -t \sin(k_z a), \Delta)^T, \quad (3.76a)$$

$$\mathbf{\Gamma} \equiv (\Gamma_0, \Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4)^T. \quad (3.76b)$$

In that form, there is a striking similarity between the minimal four-band model of a 3D topological insulator and the two-band Haldane model defined in two dimensions (see section 2.4). As a consequence, we may expect similar properties of this model with respect to the electronic band structure, the phase diagram, and the appearance of surface or edge states in appropriately chosen samples. For example, one can compute the **second Chern number** C_2 as the winding number of the unit vector $\hat{\mathbf{d}}(\mathbf{k}) \equiv \mathbf{d}(\mathbf{k})/|\mathbf{d}(\mathbf{k})|$ on the unit sphere S^4 when going around the four-dimensional Brillouin zone:

$$C_2 = \frac{3}{8\pi^2} \int d^4 k \epsilon_{\mu\nu\rho\sigma\tau} \hat{d}_\mu (\partial_{k_1} \hat{d}_\nu) (\partial_{k_2} \hat{d}_\rho) (\partial_{k_3} \hat{d}_\sigma) (\partial_{k_4} \hat{d}_\tau). \quad (3.77)$$

The above expression is well-defined as long as the bulk band gap does not vanish, *i.e.*, for $|\mathbf{d}(\mathbf{k})| \neq 0$. In the inversion-symmetric and time-reversal invariant system, this condition holds as long as $m/t \neq \pm 3$ or $m/t \neq \pm 1$. For details on the definition of the first and second Chern numbers in two and four spatial dimensions, we refer the reader to Ref. 11 and references therein.

3.6.6 Electronic band structure and bulk properties

To develop a better understanding of the minimal model for topological insulators in three dimensions [cf. Eqs. (3.72–3.75)], let us start by discussing the electronic band structure. To compute the energy eigenvalues $\mathcal{E}_\pm(\mathbf{k})$ as function of the crystal momentum \mathbf{k} we have to diagonalize the Hermitian 4×4 Bloch Hamiltonian $\mathcal{H}(\mathbf{k})$. Due to the anti-commutation properties of the Γ matrices [cf. Eq. (3.70)] we can easily read off the eigenvalues from the square of the Hamiltonian:

$$[\mathcal{H}(\mathbf{k})]^2 = [\mathbf{d}(\mathbf{k}) \cdot \mathbf{\Gamma}]^2 = |\mathbf{d}(\mathbf{k})|^2 \mathbb{1}, \quad (3.78)$$

implying that the energy bands are doubly degenerate in the presence of inversion symmetry (*i.e.*, for $\Delta'/t = 0$):

$$\mathcal{E}_\pm(\mathbf{k}) = \pm |\mathbf{d}(\mathbf{k})| = \pm \sqrt{\sum_{j=0}^4 d_j(\mathbf{k})^2}. \quad (3.79)$$

Fig. 3.15 shows the bulk band structure of a system with periodic boundary conditions in all spatial directions as function of the crystal momentum \mathbf{k} for different tuning parameters m/t .

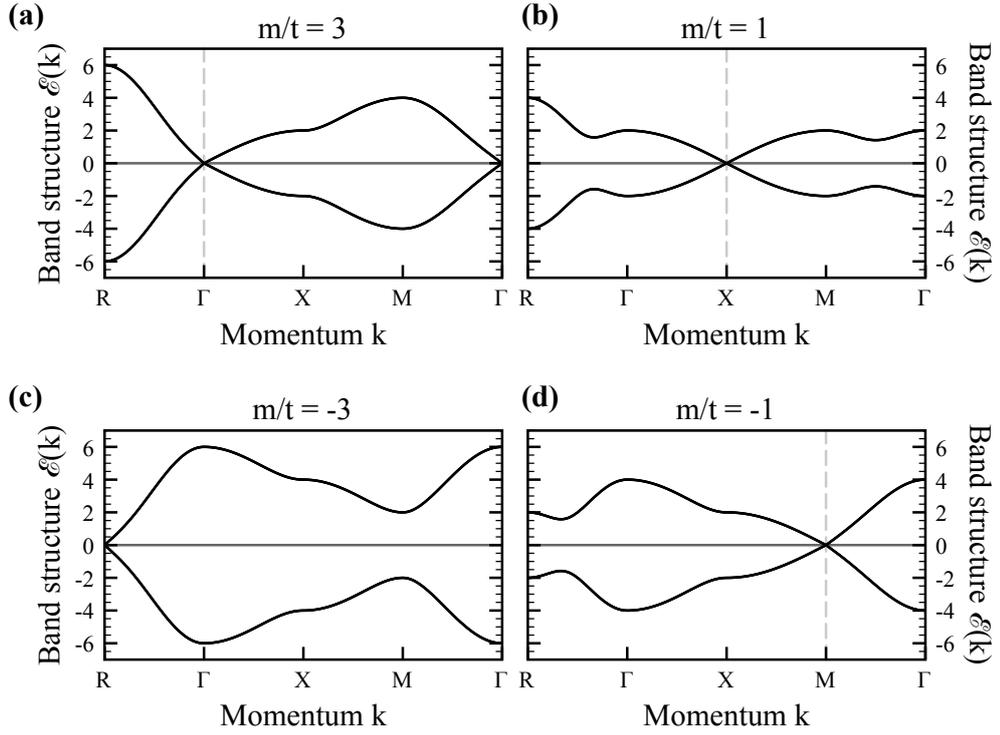


Figure 3.15. Plots of the bulk band structure $\mathcal{E}_{\pm}(\mathbf{k})$ of the time-reversal invariant and inversion-symmetric model (3.71) along the principal symmetry directions of the simple cubic lattice. Panels (a)–(d) show plots for $m/t \in \{3, 1, -1, -3\}$ and $\Delta/t = \Delta'/t = 0$ with $t = 1$. The high-symmetry points are denoted by $\Gamma = (0, 0, 0)^T$, $R = \frac{2\pi}{a}(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})^T$, $X = \frac{2\pi}{a}(\frac{1}{2}, 0, 0)^T$, and $M = \frac{2\pi}{a}(\frac{1}{2}, \frac{1}{2}, 0)^T$.

In three dimensions, the Brillouin zone of the simple cubic lattice contains eight so-called **time-reversal invariant momenta** (TRIMs) which are given by

$$\Gamma = (0, 0, 0)^T, \quad R = \frac{2\pi}{a}(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})^T, \quad (3.80a)$$

$$X_1 = \frac{2\pi}{a}(\frac{1}{2}, 0, 0)^T, \quad X_2 = \frac{2\pi}{a}(0, \frac{1}{2}, 0)^T, \quad X_3 = \frac{2\pi}{a}(0, 0, \frac{1}{2})^T, \quad (3.80b)$$

$$M_1 = \frac{2\pi}{a}(\frac{1}{2}, \frac{1}{2}, 0)^T, \quad M_2 = \frac{2\pi}{a}(0, \frac{1}{2}, \frac{1}{2})^T, \quad M_3 = \frac{2\pi}{a}(\frac{1}{2}, 0, \frac{1}{2})^T. \quad (3.80c)$$

A straightforward calculation of the spectrum of the minimal model at those eight points of high symmetry in the Brillouin zone yields:

$$\mathcal{E}_{\pm}(\Gamma + \mathbf{k}) = \pm\sqrt{(m - 3t)^2 + \Delta^2 + t^2|\mathbf{k}|^2}, \quad (3.81a)$$

$$\mathcal{E}_{\pm}(X + \mathbf{k}) = \pm\sqrt{(m - t)^2 + \Delta^2 + t^2|\mathbf{k}|^2}, \quad (3.81b)$$

$$\mathcal{E}_{\pm}(M + \mathbf{k}) = \pm\sqrt{(m + t)^2 + \Delta^2 + t^2|\mathbf{k}|^2}, \quad (3.81c)$$

$$\mathcal{E}_{\pm}(R + \mathbf{k}) = \pm\sqrt{(m + 3t)^2 + \Delta^2 + t^2|\mathbf{k}|^2}, \quad (3.81d)$$

where we have set the lattice constant $a = 1$. Apparently, for certain values of the tuning parameter m/t the conduction and valence bands show a spectrum that is reminiscent of a massive, *i.e.*, gapped Dirac theory. Furthermore, in the presence of both time-reversal and inversion symmetry, *i.e.*, for $\Delta = 0$, the electronic bulk band structure is characterized by 3D Dirac points in the bulk of the system which are located at one of the TRIMs in the Brillouin zone if the tuning

parameter m/t is chosen appropriately (see Fig. 3.15). If, however, spatial and temporal inversion symmetry are broken by, e.g., an external magnetic field, the system becomes an insulator with a band gap of $\mathcal{E}_{\text{gap}} = 2\Delta$.

To better understand the low-energy properties of the minimal model, let us focus on the case $m/t = 3$ and expand the Bloch Hamiltonian $\mathcal{H}(\mathbf{k})$ to linear order in \mathbf{k} around the Γ point:

$$\mathcal{H}(\mathbf{k}) \approx \delta m \Gamma_0 - ta(k_x \Gamma_1 + k_y \Gamma_2 + k_z \Gamma_3) + \Delta \Gamma_4 + \mathcal{O}(k^2) \quad (3.82a)$$

$$\approx \begin{bmatrix} \delta m & 0 & tak_z - i\Delta & tak_- \\ 0 & \delta m & tak_+ & -tak_z - i\Delta \\ tak_z + i\Delta & tak_- & -\delta m & 0 \\ tak_+ & -tak_z + i\Delta & 0 & -\delta m \end{bmatrix} + \mathcal{O}(k^2), \quad (3.82b)$$

where $k_{\pm} \equiv k_x \pm ik_y$, and $\delta m \equiv m - 3t$ denotes the distance to the quantum critical point. Consequently, the spectrum of this low-energy Hamiltonian describes massive Dirac fermions:

$$\mathcal{E}_{\pm}(\mathbf{k}) = \pm \sqrt{\delta m^2 + a^2 t^2 |\mathbf{k}|^2 + \Delta^2}, \quad (3.83)$$

where each band is doubly degenerate in the absence of the bulk inversion asymmetry term $\Delta' \Gamma_{04}$.

Note, however, that the two-fold degeneracy of the bulk bands is lifted in inversion-asymmetric materials such as HgTe. This can be easily understood within the minimal model upon including the inversion-symmetry breaking term $\Delta' \Gamma_{04}$ which does not commute with the Γ matrices $\Gamma_{0,\dots,4}$. In that case, the energy eigenvalues of the Bloch Hamiltonian are given by

$$\mathcal{E}_{\pm\pm}(\mathbf{k}) = \pm \sqrt{d_0(\mathbf{k})^2 + \left(\sqrt{d_1(\mathbf{k})^2 + d_2(\mathbf{k})^2 + d_3(\mathbf{k})^2} \pm d_{04}(\mathbf{k}) \right)^2 + d_4(\mathbf{k})^2} \quad (3.84)$$

with $d_{04}(\mathbf{k}) = \Delta'$. Fig. 3.16 shows the bulk band structure of the effective model with broken inversion symmetry for $m/t = 3$ and $m/t = 1$, respectively. The straightforward calculation of the bulk band structure shows that in the vicinity of the bulk Dirac point located at the center of the Brillouin zone the spectrum of the low-energy Hamiltonian is given by ($\delta m \equiv m - 3t$):

$$\mathcal{E}_{\pm\pm}(\mathbf{k}) = \pm \sqrt{\delta m^2 + \Delta^2 + (t|\mathbf{k}| \pm \Delta')^2}. \quad (3.85)$$

Thus, at the quantum critical point, $\delta m/t = \Delta/t = 0$, the bands touch at some points in momentum space not far away from the center of the Brillouin zone, Γ , but the precise location of the band touching point depends on inversion asymmetry parameter Δ'/t .

To complete the discussion of the bulk properties of the effective model for a 3D topological insulator, let us discuss the phase diagram of the system. In the presence of time-reversal symmetry, i.e., for $\Delta/t = 0$, a 3D topological insulator can be classified by four \mathbb{Z}_2 invariants $(\nu_0; \nu_1\nu_2\nu_3)$ which take on two possible values $\nu_i \in \{0, 1\}$ for $i = 0, 1, 2, 3$. As shown by Fu, Kane and Mele, ν_0 distinguishes strong topological insulators with $\nu_0 = 1$ from trivial band insulators and weak topological insulators with $\nu_0 = 0$, the latter being described by non-vanishing indices $(\nu_1\nu_2\nu_3)$.

The \mathbb{Z}_2 invariants $(\nu_0; \nu_1\nu_2\nu_3)$ for a topological insulator in three spatial dimensions can be calculated from the Bloch wave functions over the bulk Brillouin zone. The following discussion is due to Fu, Kane, and Mele^{3,8,38}. A time-reversal invariant system satisfies the relation $\hat{\Theta}\mathcal{H}(\mathbf{k})\hat{\Theta}^{-1} = \mathcal{H}(-\mathbf{k})$ for all crystal momenta \mathbf{k} within the Brillouin zone. As mentioned before, in three dimensions, there exist eight **time-reversal invariant momenta** (TRIMs) Γ_i , where

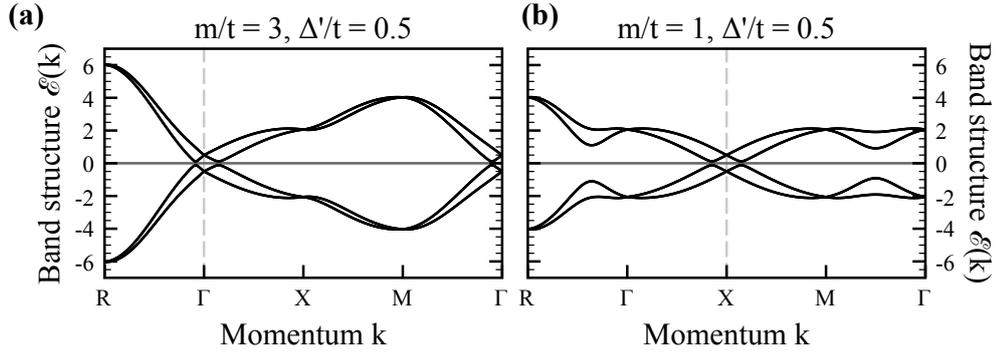


Figure 3.16. Plots of the bulk band structure $\mathcal{E}_{\pm\pm}(\mathbf{k})$ of the topological insulator model (3.71) along the principal symmetry directions of the simple cubic lattice. Panels (a) and (b) show plots for $m = 3t$ and $m = t$ with $t = 1$ and $\Delta'/t = 0.5$, respectively. Note that due to the lack of inversion symmetry, the bulk Dirac points are shifted away from the high-symmetry points denoted by $R = \frac{2\pi}{a}(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})^T$, $\Gamma = (0, 0, 0)^T$, $X = \frac{2\pi}{a}(\frac{1}{2}, 0, 0)^T$, and $M = \frac{2\pi}{a}(\frac{1}{2}, \frac{1}{2}, 0)^T$.

the Bloch Hamiltonian is invariant under time-reversal, *i.e.*, $\hat{\Theta}\mathcal{H}(\Gamma_i)\hat{\Theta}^{-1} = \mathcal{H}(\Gamma_i)$. As a consequence, the eigenstates of the Bloch Hamiltonian at those points form Kramers' pairs, *i.e.*, they are doubly degenerate. Note that the eight time-reversal invariant momenta Γ_i can also be indexed by three \mathbb{Z}_2 integers $n_l \in \{0, 1\}$, because the Γ_i are related to the reciprocal lattice vectors \mathbf{b}_1 , \mathbf{b}_2 , and \mathbf{b}_3 by

$$\Gamma_i \equiv \Gamma_{(n_1 n_2 n_3)} = \frac{1}{2} \sum_{l=1}^3 n_l \mathbf{b}_l. \quad (3.86)$$

Defining the $2N \times 2N$ unitary matrix w by

$$w_{mn}(\mathbf{k}) \equiv \langle u_m(-\mathbf{k}) | \hat{\Theta} | u_n(\mathbf{k}) \rangle, \quad (3.87)$$

where $2N$ is the number of occupied bands, and $|u_n(\mathbf{k})\rangle$ denotes the n -th Bloch state with crystal momentum \mathbf{k} , the matrix w is skew-symmetric at the time-reversal invariant momenta Γ_i , because $\langle \hat{\Theta}u | \hat{\Theta}v \rangle = \langle v | u \rangle$ and $\hat{\Theta}^2 = -\mathbb{1}$ (see also section 2.5). Mathematically, the determinant of a skew-symmetric matrix can be written as the square of the Pfaffian, a polynomial in the matrix entries, *i.e.*, the Pfaffian of the matrix w satisfies $(\text{Pf } w)^2 = \det w$. This allows for the definition time-reversal parity eigenvalues δ_i by

$$\delta_i \equiv \frac{\text{Pf } w(\Gamma_i)}{\sqrt{\det w(\Gamma_i)}} = \pm 1. \quad (3.88)$$

As shown by Fu, Kane, and Mele one can construct four \mathbb{Z}_2 invariants from the eight time-reversal invariant momenta Γ_i (Ref. 8). The first one, ν_0 , can be expressed in terms of the δ_i as a product over all eight TRIMs:

$$(-1)^{\nu_0} \equiv \prod_{i=1}^8 \delta_i = \pm 1. \quad (3.89)$$

This \mathbb{Z}_2 invariant ν_0 distinguishes between weak ($\nu_0 = 0$) and strong ($\nu_0 = 1$) 3D topological insulators which generically host an even (odd) number of Dirac states on their surfaces. The

other three invariants, ν_1 , ν_2 , and ν_3 , are given by the products of four δ_i for which the TRIMs Γ_i reside on the same plane in momentum space:

$$(-1)^{\nu_1} \equiv \prod_{n_2, n_3 \in \{0,1\}} \delta_{(n_1=1, n_2 n_3)}, \quad (3.90)$$

and ν_2 and ν_3 are obtained from the above relation by cyclic permutation of the indices. In particular, a weak topological insulator can be interpreted as being made of layers of 2D quantum spin Hall states which are stacked along the $[\nu_1 \nu_2 \nu_3]$ direction. In that sense, the indices $(\nu_1 \nu_2 \nu_3)$ are similar to Miller indices of those 2D layers. For further details on strong and weak topological insulators see, for example, Refs. 3,4,7,8,28.

In the context of our minimal model of a 3D topological insulator, we have calculated the full set of \mathbb{Z}_2 invariants $(\nu_0; \nu_1 \nu_2 \nu_3)$ as function of the tuning parameter m/t in the absence of symmetry-breaking terms, *i.e.*, for $\Delta/t = \Delta'/t = 0$. As a result, we obtain four distinct phases separated by quantum phase transitions, where the bulk band gap vanishes:

$$(\nu_0; \nu_1 \nu_2 \nu_3) = \begin{cases} (0; 000) & \text{for } |m/t| > 3 \\ (1; 111) & \text{for } -3 < m/t < -1 \\ (0; 111) & \text{for } |m/t| < 1 \\ (1; 000) & \text{for } 1 < m/t < 3 \end{cases} \quad (3.91)$$

Note that the minimal model realizes both strong and weak topological phases which in principle allows to study different topological phases within the same model. Fig. 3.17 shows the corresponding phase diagram of the effective model for a 3D topological insulator with its four different phases. In the presence of both time-reversal and inversion symmetry, the model has bulk Dirac points at the high-symmetry points in the Brillouin zone for $m/t = \pm 3$ and $m/t = \pm 1$ [cf. Eq. (3.81)]. At these points the bulk band gap vanishes, and the system undergoes a quantum phase transition from one topological state to another. For $|m/t| > 3$, the system describes a band insulator (BI) with trivial topological quantum numbers (0; 000), while for $|m/t| < 3$ the system is topologically non-trivial. To be more precise, for $1 < |m/t| < 3$ the model describes a strong topological insulator (STI) with quantum numbers (1; 000) and (1; 111) for negative and positive m , respectively. For intermediate $|m/t| < 1$, however, the model describes a weak topological insulator (WTI) which can be identified by the quantum numbers (0; 111).

3.7 Summary

In this chapter, we have discussed the electronic band structure of the so-called symmetry-protected topological insulators. In general, a topological insulator is an ordinary band insulator in the bulk of the system, where the Fermi level sits between the valence band and the conduction band, but there exist metallic states on the surfaces of the system. Importantly, those surface states traverse the band gap and allow for conduction on the surfaces. In contrast to the boundary states observed in the quantum Hall effect (see chapter 2), the surface states are protected by time-reversal symmetry and by the topologically non-trivial winding of the Bloch wave functions in the bulk band structure.

In the first part of this chapter, we have discussed the electronic properties and the topological surface states of quantum well heterostructures made of HgTe and CdTe. In particular, we have first introduced the 20-band tight-binding Hamiltonian for bulk HgTe and CdTe, both showing a zinc-blende-type lattice.

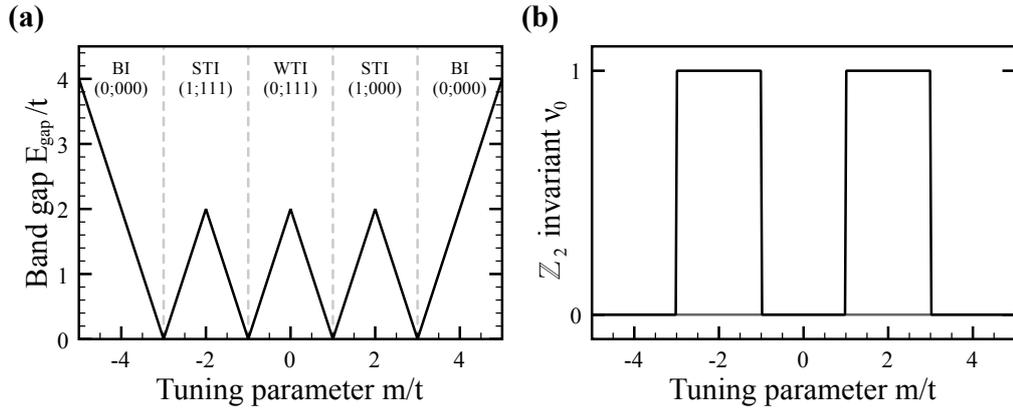


Figure 3.17. (a) Phase diagram of the minimal model for a 3D topological insulator as function of the tuning parameter m/t for $\Delta/t = \Delta'/t = 0$. At $m/t = \pm 3$ and $m/t = \pm 1$ the bulk gap vanishes, and the system undergoes a quantum phase transition between different topological quantum states which can be described in terms of four \mathbb{Z}_2 invariants ($\nu_0; \nu_1\nu_2\nu_3$). For $|m/t| > 3$, the system describes an ordinary band insulator (BI) with trivial topological quantum numbers (0; 000), while for $|m/t| < 3$ the system is topologically non-trivial. For $1 < |m/t| < 3$ the model describes a strong topological insulator (STI) with quantum numbers (1; 000) and (1; 111) for negative and positive m , respectively, and for $|m/t| < 1$ a weak topological insulator (WTI) which can be identified by the quantum numbers (0; 111). (b) Plot of the \mathbb{Z}_2 invariant ν_0 as function of the tuning parameter m , calculated numerically using the definition given by Kane and Mele³. Note that in both the trivial and the weak topological insulator phase $\nu_0 = 0$.

After that we have introduced the general terminology used in semiconductor physics to describe the bulk band structure, and we have seen that those bands are inverted in HgTe due to strong spin-orbit interactions. This has led to the theoretical prediction of the quantum spin Hall state in CdTe/HgTe/CdTe/HgTe/CdTe quantum well heterostructures, and soon after the helical edge states were experimentally found.

We also showed by explicit tight-binding calculations that the application of biaxial strain generated by the epitaxial growth of HgTe on a CdTe substrate opens up a band gap at the center of the Brillouin zone. In principle, this allows for the observation of topologically protected states on the surfaces of a strained 3D HgTe sample, because they are no longer coupled to metallic bulk states.

Finally, motivated by those results we have constructed a minimal model of strained HgTe as a 3D topological insulator, whose properties in an external magnetic field will be studied in the next chapter.

CHAPTER 4

Hall Conductance Quantization and the θ -Term in Topological Insulators

Topological insulators are new quantum states of matter which have recently attracted a great deal of theoretical and experimental interest^{3,4,7,8,28,38,40,52,53,69,71}. Unlike normal band insulators, three-dimensional topological insulators host extended metallic states on their surfaces which can be considered as massless Dirac fermions. The electronic band structure and quantum spin texture of the helical surface states have been well established theoretically and experimentally, as discussed in chapters 2 and 3.

Particularly interesting is the problem of a 3D topological insulator subjected to an external magnetic field. Since the 2D surface Dirac fermions carry definite charge $-e$, the magnetic field couples to the orbital motion of the Dirac fermions. If the magnetic field contains a component normal to the two-dimensional surfaces, this leads to a Landau level quantization of the electronic surface band structure. Landau levels for Dirac fermions are special, however, because a Landau level at exactly zero energy is guaranteed to exist^{95,96}. This zeroth Landau level is particle-hole symmetric in the sense that the Hall conductance is equal and has opposite sign when the Landau level is either empty or fully occupied. Since the Hall conductance increases by a full conductance quantum e^2/h when the Fermi level crosses a Landau level, the Hall conductance of the surface states σ_{xy} must be half-integer quantized^{11,28}. Based on this unique electric response to an applied magnetic field, Qi *et al.* have proposed the unconventional **magneto-electric effect** which is regarded as one of the characteristic features of strong topological insulators^{11,12}. However, in a usual quantum Hall system, the current-carrying 1D edge states are responsible for the integer-quantized Hall conductance σ_{xy} measured in transport experiments (see chapter 2). It is, however, not immediately clear if similar boundary states are also responsible for the half-quantized Hall conductance in strong topological insulators, and how the quantized nature of the conducting edge channels can be reconciled with the prediction of the half quantization of the Hall conductance.

Experimentally, quantum oscillations originating from the Landau level quantization of the surface states have been observed in several transport experiments^{10,97-100}. For example, in a recent experiment, Brüne *et al.* have reported the observation of the quantum Hall effect in a strained bulk mercury telluride (HgTe) samples¹⁰. Notably, quantum Hall plateaus seem to appear at various magnetic field strengths. In contrast to the quantum Hall effect of an ordinary 2D electron gas, they report an odd-integer sequence of Hall plateaus with filling factors $\nu = 9, 7, 5$ at low magnetic fields, before the Hall plateaus continue in the usual sequence with filling factors $\nu = 4, 3, 2$ at high magnetic fields (see Fig. 4.1). The occurrence of the odd-integer filling

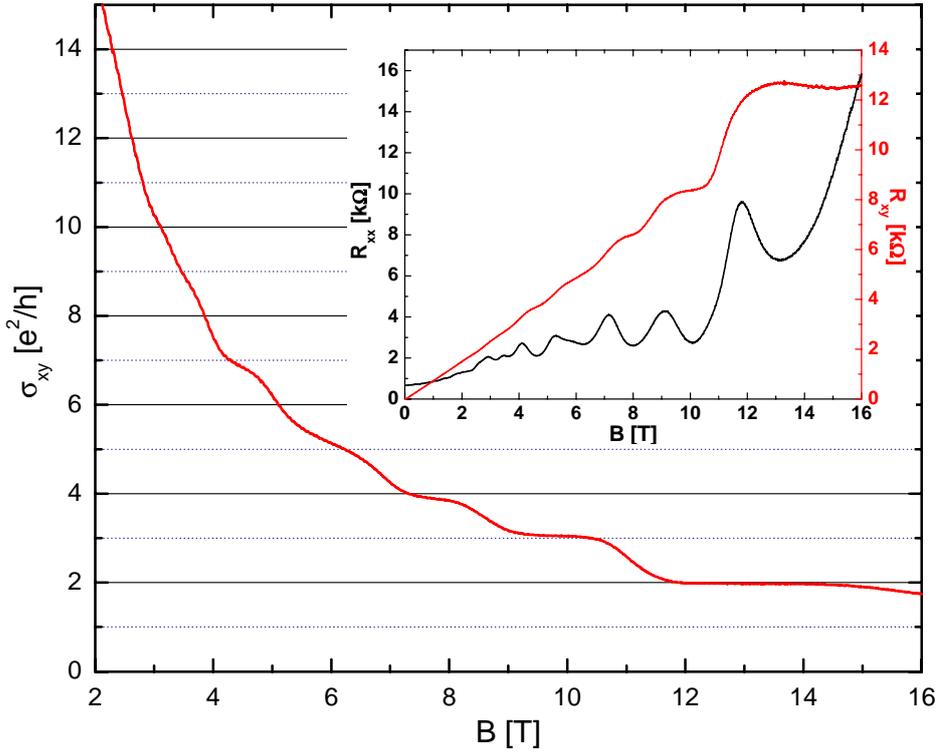


Figure 4.1. Transport data of a strained 70 nm thick HgTe sample measured at 50 mK, as obtained by Brüne *et al.* (Ref. 10). The Hall conductivity (red line) shows plateaus at the integers $\nu = 9, 7, 5$ and at $\nu = 4, 3, 2$. The inset shows the Hall resistance R_{xy} (red line) and the longitudinal resistance R_{xx} (black line). The Hall resistance shows plateaus at the same magnetic fields, where the longitudinal resistance R_{xx} develops minima. Figure after Ref. 10.

factors is attributed to the presence of a Landau level at zero energy. This is but the smoking gun for the quantum Hall of Dirac fermions on the 2D surfaces of the sample, because massless Dirac fermions subjected to an orbital magnetic field always exhibit a zero mode, as shown below. However, at the same time plateaus develop, the longitudinal resistance ρ_{xx} does not vanish (see Fig. 4.1) which indicates that the conductance measurements are either contaminated by metallic states present on the side surfaces or by finite bulk conductivity. As pointed out by Brüne *et al.*, a likely candidate for the finite longitudinal resistance in HgTe are the metallic side surfaces: Due to the particular orientation of the magnetic field with respect to the sample, the side surfaces are subjected to a parallel magnetic field rather than a perpendicular magnetic field. Thus, the surface states on the side surfaces remain metallic and coexist with the Landau levels formed by the Dirac fermions in the surfaces normal to the magnetic field, likely giving rise to the non-vanishing longitudinal resistance.

In addition, the spin-orbit interaction, which actually is responsible for the appearance of the Dirac fermions in HgTe in the first place, makes the Zeeman coupling different from that in graphene which also exhibits Dirac fermions at low energies. Instead of a simple spin-splitting of the electronic band structure, the Zeeman interaction coupling the electron's spin to the magnetic field acts as a Dirac mass term in topological insulators. As we show below, the interplay between the orbital magnetic field, the uniform spin-splitting Zeeman field, and the inversion asymmetry present in zinc-blende-type crystal structures such as HgTe leads to a complicated interplay of the associated energy gaps. As a consequence, the Dirac electrons located on

different surfaces show different behavior in presence of a magnetic field. Moreover, since all surface states of a finite three-dimensional sample of a topological insulator are connected, the surface states living on the sample surfaces may, in principle, move from one side to another.

To address those issues, we investigate the Hall response of a 3D strong topological insulator in the presence of a uniform magnetic field oriented such that all surfaces have a normal component, thus opening a gap for all surface excitations. Although a uniform magnetic field breaks time-reversal and translational symmetries, we can calculate the electronic band structure of a tight-binding model by choosing a suitable sample geometry and gauge of the vector potential, as discussed in chapter 2. Starting from the minimal model for strained 3D HgTe derived in the previous chapter we compute the effective **surface Dirac Hamiltonian**, describing massive Dirac fermions in the presence of an orbital magnetic field, using the standard bound state solution of an interface between a topological and a trivial band insulator. We then consider the chiral edge states which appear in finite magnetic fields, and discuss their dependence on model parameters, in particular the Dirac mass term. For a sample with finite dimensions in all directions, those edge states connect and form **edge state networks**, whose properties will be discussed as well. After that, we clarify the role of the so-called **θ -term** in topological insulators and show how it is related to the Hall response σ_{xy} measured in experiments. Notably, we show that the θ -term can be interpreted as a local measurement of the Hall conductance $\sigma_{xy}^{\text{surf}}$ for a single surface, whereas a transport experiment to determine $\sigma_{xy}^{\text{Hall}}$ is but a global measurement of the whole sample. We show that the surface Hall conductance and the θ -term are related by $\sigma_{xy}^{\text{surf}} = (\theta/2\pi) \times (e^2/h)$ for a single surface of a strong topological insulator, while $\sigma_{xy}^{\text{Hall}} = n(e^2/h)$ with integer n in transport measurements, and we derive the general relation between those two quantities. Finally, we study the role of the edge channels and the θ -term in the presence of an external magnetic field by means of exact diagonalization techniques and compare the numerical results with our obtained analytical results.

4.1 Quantum Hall states in the minimal model for strained 3D HgTe

To develop a better understanding of the quantum Hall effect in a topological insulator, we consider the minimal model which was derived from strained 3D HgTe in the previous chapter. After a brief recap of that model, we derive the surface Dirac Hamiltonian and compute the Landau levels sequence. Here, we observe that the Zeeman interaction coupling the spin of the Dirac fermions to the magnetic field acts as a Dirac mass term, thereby modifying the Landau level sequence.

4.1.1 Minimal model for topological insulators

In the previous chapter, we have shown that strained 3D HgTe can, in principle, host Dirac-like states on its 2D surfaces. Here, note that the application of biaxial strain due to growth on a CdTe substrate opens up a band gap which decouples the surface states from the otherwise metallic bulk bands. As discussed before, transport measurements by Brüne *et al.* have provided some evidence for the strong topological insulator phase of strained 3D HgTe and the Dirac-like surface states¹⁰.

To be specific, we study the following tight-binding Hamiltonian on a simple cubic lattice:

$$H_0 = m \sum_{\mathbf{R}} \Psi_{\mathbf{R}}^\dagger \Gamma_0 \Psi_{\mathbf{R}} - t \sum_{\mathbf{R}} \sum_{j=1}^3 \left[\Psi_{\mathbf{R}}^\dagger \left(\frac{\Gamma_0 - i\Gamma_j}{2} \right) \Psi_{\mathbf{R}-\mathbf{e}_j} + \text{H.c.} \right], \quad (4.1)$$

where m is the so-called **tuning parameter**, and t denotes the **overlap parameter** or **hopping amplitude** between nearest-neighboring sites. Moreover, the Γ matrices are constructed as follows:

$$\{\Gamma_0, \Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4\} \equiv \{\tau_z \otimes \sigma_0, -\tau_x \otimes \sigma_x, -\tau_x \otimes \sigma_y, -\tau_x \otimes \sigma_z, \tau_y \otimes \sigma_0\}. \quad (4.2)$$

Here, $\boldsymbol{\tau} = (\tau_x, \tau_y, \tau_z)^T$ is the vector of Pauli matrices acting on the orbital degrees of freedom, and $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)^T$ denotes the usual Pauli matrices acting on the spin degree of freedom. By Fourier decomposition of the four-component spinors $\Psi_{\mathbf{R}}$,

$$\Psi_{\mathbf{R}} = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{R}} \Psi_{\mathbf{k}}, \quad \Psi_{\mathbf{k}} = \frac{1}{\sqrt{N}} \sum_{\mathbf{R}} e^{-i\mathbf{k} \cdot \mathbf{R}} \Psi_{\mathbf{R}}, \quad (4.3)$$

where N denotes the number of crystallographic unit cells, we obtain the corresponding Bloch Hamiltonian of the 3D topological insulator:

$$\mathcal{H}_0(\mathbf{k}) = m \Gamma_0 - t \sum_{j=1}^3 (\cos(k_j a) \Gamma_0 + \sin(k_j a) \Gamma_j). \quad (4.4)$$

We also consider additional terms in the Hamiltonian which preserve the rotational symmetries of the minimal model, but break time-reversal and/or inversion symmetry. As shown in the previous chapter, we find two terms: The first term, $\Delta \Gamma_4$, breaks both time-reversal and inversion symmetry, whereas the second term, $\Delta' \Gamma_{04}$, will be present in any system without inversion center such as HgTe. Note that the application of an external magnetic field breaks time-reversal and inversion symmetry, thus it will induce a small, but finite contribution to $\Delta \Gamma_4$. Since $\Delta \Gamma_4$ breaks both time-reversal and inversion symmetry, it usually suffices to consider only that term in our discussion of the Hall response and the θ -term. Hence, in the following, we usually neglect the effect of $\Delta' \Gamma_{04}$ on the electronic band structure as it does not change qualitatively our results. Moreover, since both terms are diagonal in both momentum space and real space, one can simply add them to the corresponding real-space or momentum-space Hamiltonian if needed.

The minimal model for 3D topological insulators exhibits four distinct phases which are separated by quantum phase transitions, where the bulk band gap vanishes, as described in section 3.6.6. Fig. 4.2 shows the corresponding phase diagram of the minimal model with the four different phases. In the following, we focus on the strong topological insulator phase for $1 < m/t < 3$, as indicated by the red shading in Fig. 4.2.

4.1.2 Effect of an applied external magnetic field

The problem of a 3D topological insulator subjected to an external magnetic field is surprisingly rich. On the one hand, the magnetic field couples to the orbital motion of the 2D surface Dirac fermions which carry charge $-e$, leading to the appearance of Landau levels when the magnetic field contains a component normal to the surface. Note that the magnetic field component parallel to the surface can be accounted for by a suitable shift of the momenta \mathbf{k} in the Hamiltonian. On the other hand, the magnetic field also couples to the electronic spin of the Dirac fermions, leading to a Zeeman energy splitting of spin-up and spin-down states. However, due to the spin-orbit interaction in HgTe, the effect of the Zeeman interaction on the electronic band structure is quite different from the one observed for usual Dirac fermions which are found, for example, in graphene. As we show below, the magnetic Zeeman field acts as a Dirac mass term

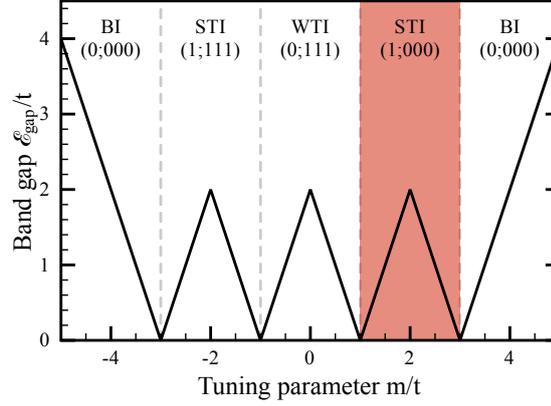


Figure 4.2. Phase diagram of the minimal model for a 3D topological insulator as function of the tuning parameter m . At $m/t = \pm 3$ and $m/t = \pm 1$ the bulk gap vanishes, and the system undergoes a quantum phase transition between different topological quantum states which can be described in terms of four \mathbb{Z}_2 invariants $(\nu_0; \nu_1 \nu_2 \nu_3)$. For $|m/t| > 3$, the system describes an ordinary band insulator (BI) with trivial topological quantum numbers $(0; 000)$, while for $|m/t| < 3$ the system is topologically non-trivial. For $1 < |m/t| < 3$ the model describes a strong topological insulator (STI) with quantum numbers $(1; 000)$ and $(1; 111)$ for negative and positive m , respectively, and for $|m/t| < 1$ a weak topological insulator (WTI) which can be identified by the quantum numbers $(0; 111)$. In this chapter, we focus on the strong topological insulator phase for $1 < m/t < 3$, as indicated by the red shading.

in topological insulators instead of a simple spin-splitting of the electronic band structure. To study topological insulators in an applied magnetic field and investigate the Hall response, we briefly discuss how the magnetic field can be included into the minimal model of a 3D topological insulator in terms of (i) a Zeeman interaction coupling the electronic spin to the magnetic field and (ii) an orbital magnetic field coupling the charge to the magnetic field.

4.1.2.1 The magnetic Zeeman interaction

In the presence of an applied external magnetic field, the two-fold degeneracy of the bulk bands may also be lifted by the magnetic **Zeeman interaction** $\mathcal{H}_Z = g\mu_B \mathbf{B}_Z \cdot \boldsymbol{\sigma}$ through a spin-dependent Zeeman energy. Here, the constant of proportionality g is usually referred to as **g -factor**, and μ_B denotes the **Bohr magneton**. Both the electron (E) and the light-hole (LH) band are subject to the Zeeman spin-splitting, and we can represent the above Zeeman Hamiltonian in the basis set of Eq. (3.50) as:

$$\mathcal{H}_Z = g_E \mu_B \left(\frac{\tau_0 + \tau_z}{2} \right) \otimes (\mathbf{B}_Z \cdot \boldsymbol{\sigma}) + g_{LH} \mu_B \left(\frac{\tau_0 - \tau_z}{2} \right) \otimes (\mathbf{B}_Z \cdot \boldsymbol{\sigma}) \quad (4.5a)$$

$$= \mu_B \begin{bmatrix} g_E B_Z^z & g_E B_Z^- & 0 & 0 \\ g_E B_Z^+ & -g_E B_Z^z & 0 & 0 \\ 0 & 0 & g_{LH} B_Z^z & g_{LH} B_Z^- \\ 0 & 0 & g_{LH} B_Z^+ & -g_{LH} B_Z^z \end{bmatrix}, \quad (4.5b)$$

where g_E and g_{LH} denote the g -factors of the electron and light-hole bands, respectively, and \mathbf{B}_Z denotes the Zeeman field. Furthermore, we have introduced $B_Z^\pm \equiv B_Z^x \pm iB_Z^y$. In terms of the matrices Γ_a [cf. Eq. (4.2)] and their ten commutators Γ_{ab} , the Zeeman Hamiltonian can also

be rewritten as

$$\begin{aligned}\mathcal{H}_Z = & \mu_B B_Z^x (g_+ \Gamma_{23} - g_- \Gamma_{14}) \\ & - \mu_B B_Z^y (g_+ \Gamma_{13} + g_- \Gamma_{24}) \\ & + \mu_B B_Z^z (g_+ \Gamma_{12} - g_- \Gamma_{34}),\end{aligned}\quad (4.6)$$

where, for brevity, we have introduced $g_{\pm} \equiv (g_E \pm g_{LH})/2$.

4.1.2.2 The orbital magnetic field

As already discussed in chapter 2, in the presence of a magnetic field $\mathbf{B}(\mathbf{r})$ the kinetic momentum \mathbf{p} must be replaced by the canonical momentum operator $\Pi \equiv \mathbf{p} - q\mathbf{A}(\mathbf{r})$, where $\mathbf{A}(\mathbf{r})$ is the vector potential of the magnetic field. On a discrete lattice, however, the analogue of **minimal substitution** is the so-called **Peierls substitution** in which the overlap parameter t is augmented by an additional phase factor¹⁰¹:

$$t \rightarrow t' = t \exp\left(i \frac{e}{\hbar c} \int_{\mathbf{r}_i}^{\mathbf{r}_f} d\mathbf{r} \cdot \mathbf{A}(\mathbf{r})\right), \quad (4.7)$$

where $\mathbf{r}_{i,f}$ denote the initial and final position of the electron. To be specific, let us consider a uniform magnetic field in the yz plane given by the magnetic field vector

$$\mathbf{B}(\mathbf{r}) = B_0 (0, \cos \phi, \sin \phi)^T, \quad (4.8)$$

where ϕ describes the inclination angle of the magnetic field with respect to the y axis. The corresponding Landau gauge points in x direction, but varies only in y and z directions:

$$\mathbf{A}(\mathbf{r}) = B_0 (z \cos \phi - y \sin \phi) \mathbf{e}_x = (zB_y - yB_z) \mathbf{e}_x. \quad (4.9)$$

As a consequence, the overlap parameter for electrons hopping in x direction acquires a position-dependent phase factor given by

$$t'_x = t \exp\left[\frac{ie}{\hbar c} (zB_y - yB_z)\right] = t \exp[2\pi i (zN_{\Phi}^y - yN_{\Phi}^z)], \quad (4.10)$$

while the hopping amplitudes in y and z direction remain unaffected, $t'_y = t'_z = t$. For brevity, we have introduced the **magnetic flux densities** as ($\alpha \in \{y, z\}$):

$$N_{\Phi}^{\alpha} \equiv \frac{\Phi_{\alpha}}{\Phi_0} = \frac{eB_{\alpha}}{\hbar c}. \quad (4.11)$$

Note that the resulting tight-binding Hamiltonian explicitly depends on the Cartesian coordinates y and z which spoils the translational invariance in those directions, but the translational invariance in x direction prevails. Hence, k_x is a good quantum number, and we may perform a Fourier decomposition of the four-component spinors $\Psi_{\mathbf{R}} = \Psi_{xyz}$ with respect to the x direction, as discussed in chapter 2. This leads to the following two-dimensional tight-binding Hamiltonian, where the magnetic field enters through the flux-dependent shift of momentum (second and third term) and through the Zeeman interaction coupling the magnetic field and

the spin degree of freedom (last term):

$$\begin{aligned}
\mathcal{H}(k_x) = & \sum_{y,z} \Psi_{k_x y z}^\dagger (m \Gamma_0 + \Delta \Gamma_4) \Psi_{k_x y z} \\
& - t \sum_{y,z} \Psi_{k_x y z}^\dagger \cos(k_x a + 2\pi(zN_\Phi^y - yN_\Phi^z)) \Gamma_0 \Psi_{k_x y z} \\
& - t \sum_{y,z} \Psi_{k_x y z}^\dagger \sin(k_x a + 2\pi(zN_\Phi^y - yN_\Phi^z)) \Gamma_1 \Psi_{k_x y z} \\
& - t \sum_{y,z} \left[\Psi_{k_x y z}^\dagger \left(\frac{\Gamma_0 - i\Gamma_2}{2} \right) \Psi_{k_x, y-a, z} + \text{H.c.} \right] \\
& - t \sum_{y,z} \left[\Psi_{k_x y z}^\dagger \left(\frac{\Gamma_0 - i\Gamma_3}{2} \right) \Psi_{k_x, y, z-a} + \text{H.c.} \right] + \mathcal{H}_Z.
\end{aligned} \tag{4.12}$$

Here, we have included the time-reversal and inversion symmetry-breaking term $\Delta \Gamma_4$ which is induced by a finite magnetic field and turns out to be important for the Landau level sequence, as we discuss later. Due to the explicit dependence of the Hamiltonian on the spatial coordinates in y and z direction, we can only calculate the electronic band structure of samples in a quasi-1D geometry, for example, in a beam-shape geometry with periodic boundary conditions in x direction and open boundary conditions in y and z directions. We have used this particular tight-binding Hamiltonian for our numerical calculations of the electronic band structure and the so-called θ -term introduced later, and the results shown in the last section are in very good agreement (up to a few percent due to finite-size effects) with the following analytical approach.

4.1.3 Derivation of the surface Dirac Hamiltonian

Before we discuss the numerical results, let us first consider an analytical approach to the band structure of the minimal model in the presence of an external magnetic field. To show the appearance of topologically non-trivial states on the 2D surfaces of a 3D topological insulator, we consider an interface between a strong topological insulator and a trivial band insulator with normal vector \hat{n} [see Fig. 4.3 (a)]. Starting from the bulk Bloch Hamiltonian of our time-reversal invariant and inversion-symmetric model of a 3D topological insulator,

$$\mathcal{H}(\mathbf{k}) = m \Gamma_0 - t \sum_{j=1}^3 (\cos(k_j a) \Gamma_0 + \sin(k_j a) \Gamma_j), \tag{4.13}$$

we introduce new coordinates such that $\mathbf{r} = \mathbf{r}_\parallel + \mathbf{r}_\perp$, where $\mathbf{r}_\perp \equiv (\mathbf{r} \cdot \hat{n}) \hat{n} \equiv r_\perp \hat{n}$ denotes the out-of-plane component of \mathbf{r} , while \mathbf{r}_\parallel describes the in-plane component of \mathbf{r} parallel to the interface. Since the phases on both sides of the interface are characterized by different tuning parameters $m/t < 3$ and $m/t > 3$, respectively, let us consider the effective mass parameter $\delta m(r_\perp) \equiv m(r_\perp) - 3t$ which changes sign at the interface as function of the out-of-plane position r_\perp [see Fig. 4.3 (b)]. Throughout this thesis, we use the convention that the surface normal vector points away from the topological insulator, so that $\delta m(r_\perp) > 0$ for $r_\perp > 0$ mimics the trivial band insulator, while $\delta m(r_\perp) < 0$ for $r_\perp < 0$ describes the strong topological insulator.

To compute the low-energy 2D Dirac Hamiltonian describing the topological surface states at the interface, we first compute the eigenmodes of the linearized bulk Hamiltonian and then project this Hamiltonian onto the eigenstates of its zero-energy eigenmode, as shown below

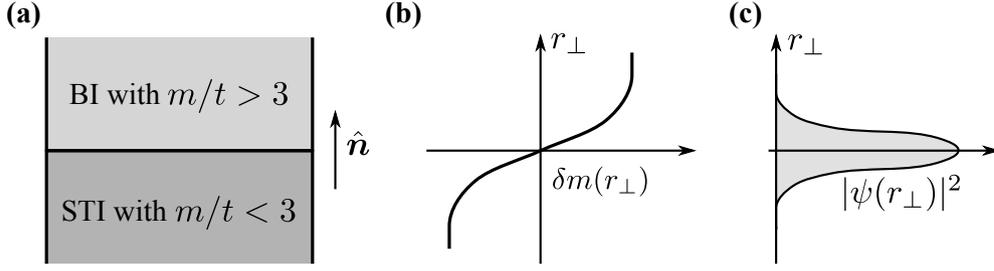


Figure 4.3. (a) Schematic picture of an interface between a strong topological insulator (STI) and a trivial band insulator (BI). (b) The effective mass parameter $\delta m(r_\perp) \equiv m(r_\perp) - 3t$ changes sign across the interface as function of the normal coordinate r_\perp . Here, $\delta m(r_\perp) > 0$ for $r_\perp > 0$ mimics the trivial band insulator, while $\delta m(r_\perp < 0)$ for $r_\perp < 0$ describes the strong topological insulator. (c) Plot of the spatial probability amplitude of the surface state, $|\psi(r_\perp)|^2$, as function of r_\perp .

[see Fig. 4.3 (c)]. Close to the interface, *i.e.*, for $|\delta m(r_\perp)| = |m(r_\perp) - 3t| \ll t$, we may perform a straightforward Taylor expansion of the bulk Bloch Hamiltonian to linear order in \mathbf{k} around the Γ point. As a result, we obtain the low-energy bulk Dirac Hamiltonian as follows:

$$\mathcal{H}(\mathbf{k}) \approx \delta m(r_\perp)\Gamma_0 - ta \sum_{j=1}^3 k_j \Gamma_j = \delta m(r_\perp)\Gamma_0 - v_F \mathbf{k} \cdot \tilde{\Gamma}, \quad (4.14)$$

where, for brevity, we have introduced $\tilde{\Gamma} \equiv (\Gamma_1, \Gamma_2, \Gamma_3)^T$, and $v_F \equiv ta$ denotes the bulk Fermi velocity. Similar to the decomposition of the real space coordinate \mathbf{r} , we can split up the crystal momentum into two contributions \mathbf{k}_\parallel and \mathbf{k}_\perp , where \mathbf{k}_\parallel and $\mathbf{k}_\perp \equiv (\mathbf{k} \cdot \hat{\mathbf{n}})\hat{\mathbf{n}}$ denote the in-plane and out-of-plane momenta, respectively¹. Due to the interface, the translational invariance is broken in the normal direction. As a consequence, upon substituting $\mathbf{k}_\perp \rightarrow -i\hbar\nabla_{r_\perp} = -i\hbar\hat{\mathbf{n}}\partial_{r_\perp}$ we obtain a one-dimensional Schrödinger equation for the surface state with in-plane momentum \mathbf{k}_\parallel :

$$\mathcal{H}(\mathbf{k}_\parallel, r_\perp)\Psi(\mathbf{k}_\parallel, r_\perp) \doteq \mathcal{E}(\mathbf{k}_\parallel)\Psi(\mathbf{k}_\parallel, r_\perp), \quad (4.15)$$

where the one-dimensional Bloch Hamiltonian $\mathcal{H}(\mathbf{k}_\parallel, r_\perp)$ is given by

$$\mathcal{H}(\mathbf{k}_\parallel, r_\perp) = \delta m(r_\perp)\Gamma_0 - v_F \mathbf{k}_\parallel \cdot \tilde{\Gamma} + i\hbar v_F \hat{\mathbf{n}} \cdot \tilde{\Gamma} \partial_{r_\perp}. \quad (4.16)$$

Due to the assumed translational invariance parallel to the surface, we can make a plane-wave ansatz for the wave function $\Psi(\mathbf{k}_\parallel, r_\perp) \propto \exp(i\mathbf{k}_\parallel \cdot \mathbf{r}_\parallel)\psi(r_\perp)$. Substituting this ansatz into the one-dimensional Schrödinger equation for the full wave function [cf. Eq. (4.15)], the topological surface states are given by the zero-energy solutions of the following one-dimensional Schrödinger equation for $\psi(r_\perp)$:

$$H_\perp \psi(r_\perp) = (\delta m(r_\perp)\Gamma_0 + i\hbar v_F \hat{\mathbf{n}} \cdot \tilde{\Gamma} \partial_{r_\perp})\psi(r_\perp) \doteq 0. \quad (4.17)$$

This Schrödinger equation is but a homogeneous linear one-dimensional ordinary differential equation for $\psi(r_\perp)$:

$$\partial_{r_\perp} \psi(r_\perp) = \frac{i}{\hbar v_F} (\hat{\mathbf{n}} \cdot \tilde{\Gamma})^{-1} \Gamma_0 \delta m(r_\perp) \psi(r_\perp). \quad (4.18)$$

¹Different conventions for the definition of the in-plane momentum \mathbf{k}_\parallel appear in the literature. On the one hand, a common choice is $\mathbf{k}_\parallel = (k_x, k_y)^T$ which results in an effective Dirac Hamiltonian of the usual form, $v_F(k_x\sigma_x + k_y\sigma_y)$. On the other hand, if one defines $\mathbf{k}_\parallel \equiv \hat{\mathbf{n}} \times \mathbf{k}$, then the resultant Dirac Hamiltonian takes the form $v_F(k_x\sigma_y - k_y\sigma_x)$ for $\hat{\mathbf{n}} = \mathbf{e}_z$. Up to a rotation about the z axis, both conventions are, however, equivalent.

A general solution of that differential equation can be obtained by choosing a suitable ansatz for the wave function $\psi(r_\perp)$:

$$\psi(r_\perp) = \psi_0 e^{-f(r_\perp)}. \quad (4.19)$$

Here, ψ_0 is taken to be an eigenvector of the Hermitian matrix $\Lambda \equiv -i(\hat{\mathbf{n}} \cdot \tilde{\Gamma})^{-1} \Gamma_0$ with eigenvalue λ , i.e., $\Lambda \psi_0 = \lambda \psi_0$. By substituting this ansatz into the differential equation (4.17) we then obtain a simple linear ordinary differential equation for $f(r_\perp)$:

$$\partial_{r_\perp} f(r_\perp) = \lambda \frac{\delta m(r_\perp)}{\hbar v_F}. \quad (4.20)$$

As a result, the straightforward integration with respect to r_\perp yields a generic solution for the surface state $\psi(r_\perp)$:

$$\psi(r_\perp) = \psi_0 \exp\left(-\lambda \int^{r_\perp} dr' \frac{\delta m(r')}{\hbar v_F}\right). \quad (4.21)$$

Since $\delta m(r_\perp) > 0$ for $r_\perp > 0$, a physical solution is given by a positive eigenvalue λ . Note that the eigenvalues and eigenvectors of the matrix Λ can be obtained by analytical diagonalization, leading to four normalized eigenvectors $\hat{\mathbf{v}}_{\pm,1}$ and $\hat{\mathbf{v}}_{\pm,2}$ corresponding to the eigenvalues $\lambda = \pm 1$:

$$\hat{\mathbf{v}}_{\pm,1} = \frac{1}{\sqrt{2}} (\mp i n_z, \mp i n_+, 1, 0)^T, \quad \hat{\mathbf{v}}_{\pm,2} = \frac{1}{\sqrt{2}} (\mp i n_-, \pm i n_z, 0, 1)^T. \quad (4.22)$$

Thus, by projecting the bulk Dirac Hamiltonian (4.14) onto the eigenmodes $\mathbf{v}_{+,1}$ and $\mathbf{v}_{+,2}$ corresponding to the eigenvalue $\lambda = 1$, we obtain the proper in-plane Hamiltonian $\overline{\mathcal{H}}_0$. For the sake of concreteness, we choose $\hat{\mathbf{n}} = \mathbf{e}_z$ in the following, but the results can be easily generalized to an arbitrary surface normal $\hat{\mathbf{n}}$. After a unitary transformation given by $\mathcal{U}_z = \exp(i\pi\sigma_z/4)$, the resulting surface Hamiltonian $\overline{\mathcal{H}}_0$ takes the following form:

$$\overline{\mathcal{H}}_0(\mathbf{k}_\parallel) = v_F \mathbf{k} \cdot \boldsymbol{\sigma}^*, \quad (4.23)$$

where z^* denotes the complex conjugate of the complex number z . In a similar way, we can project the additional term $\Delta \Gamma_4$, which simultaneously breaks time-reversal symmetry and inversion symmetry of the 3D topological insulator, onto the physical surface eigenstates $\mathbf{v}_{+,1}$ and $\mathbf{v}_{+,2}$ as well, and we obtain:

$$\overline{\mathcal{H}}' = -\Delta \sigma_z. \quad (4.24)$$

In other words, the symmetry-breaking term $\propto \Delta$ acts as a Dirac mass term with constant magnitude. We can also project the magnetic Zeeman interaction \mathcal{H}_Z [cf. Eq. (4.5)] onto the eigenmodes of the surface state. After a proper shift of the in-plane momentum \mathbf{k}_\parallel to absorb Zeeman terms proportional to σ_x and σ_y , the Zeeman interaction results in a second mass term for the Dirac fermions. In contrast to $\overline{\mathcal{H}}'$, however, this mass term depends explicitly on the orientation of the surface under discussion:

$$\overline{\mathcal{H}}_Z = [(g_- B_z^x, g_- B_z^y, -g_+ B_z^z)^T \cdot \hat{\mathbf{n}}] \sigma_z. \quad (4.25)$$

Finally, the orbital magnetic field coupling to the electronic charge is taken into account by minimal substitution, $\mathbf{k} \rightarrow \mathbf{k} + \frac{e}{c} \mathbf{A}(\mathbf{r})$, where $\mathbf{A}(\mathbf{r})$ is the vector potential of the normal field component appropriately transformed into the proper basis with respect to the surface. As a result, for small $|\delta m/t|$ and for $|\Delta|, |B_z^\alpha|, |\Phi_\alpha| \ll |\delta m|$ we obtain the following **surface Dirac Hamiltonian**:

$$\mathcal{H}_{\text{surf}}(\mathbf{k}) = v_F \left(\mathbf{k} + \frac{e}{c} \mathbf{A}(\mathbf{r}) \right) \cdot \boldsymbol{\sigma}^* + m_{\text{surf}} \sigma_z, \quad (4.26)$$

which describes massive Dirac fermions on the two-dimensional surface of a strong topological insulator with in-plane momentum \mathbf{k} and subjected to a perpendicular orbital magnetic field that is given by $\mathbf{A}(\mathbf{r})$. It is worth to note that the unusual Dirac mass term

$$m_{\text{surf}} \equiv (g_- B_z^x, g_- B_z^y, -g_+ B_z^z)^T \cdot \hat{\mathbf{n}} - \Delta \quad (4.27)$$

depends on the relative orientation of the magnetic field with respect to the surface normal vector $\hat{\mathbf{n}}$. In principle, this allows us to independently control the effects due to the orbital magnetic field and the Zeeman magnetic field. For example, the Dirac mass m_{surf} can be finite even for a magnetic field parallel to the surface for which the vector potential $\mathbf{A}(\mathbf{r})$ vanishes, as discussed below.

4.1.4 Landau level spectrum of the surfaces

To compute the Landau level spectrum of the massive Dirac Hamiltonian (4.26) in the presence of an external orbital magnetic field $\mathbf{B}(\mathbf{r}) = \text{curl } \mathbf{A}(\mathbf{r})$, we first consider massless Dirac fermions moving in the two-dimensional xy plane subjected to a perpendicular magnetic field $\mathbf{B} = B_\perp \mathbf{e}_z$. Later, we include a finite Dirac mass term which leads to a small correction of the Landau levels and in particular shifts the zeroth Landau level away from zero energy.

Let us start with a discussion of the massless Dirac Hamiltonian [cf. Eq. (4.26)]

$$\mathcal{H}_0(\mathbf{k}) = v_F \left(\mathbf{k} + \frac{e}{c} \mathbf{A}(\mathbf{r}) \right) \cdot \boldsymbol{\sigma}^*, \quad (4.28)$$

where v_F is the velocity of the Dirac fermions, and $\boldsymbol{\sigma}^* = (\sigma_x, -\sigma_y)^T$ is a vector of Pauli matrices in the 2D xy plane. A convenient choice for the vector potential $\mathbf{A}(\mathbf{r})$ of the perpendicular magnetic field $\mathbf{B}(\mathbf{r}) = B_\perp \mathbf{e}_z$ is the **Landau gauge**:

$$\mathbf{B}(\mathbf{r}) = B_\perp \mathbf{e}_z, \quad \mathbf{A}(\mathbf{r}) = -y B_\perp \mathbf{e}_x, \quad (4.29)$$

which points in x direction, but varies with the y coordinate in real space. Within the Landau gauge, the system possesses translational invariance in the x direction, motivating the following ansatz for the wave function of a Dirac fermion:

$$\Psi_{k_x}(y) \propto e^{ik_x x} \psi(y), \quad (4.30)$$

where k_x denotes the conserved momentum along the x direction. As a result of this wave function ansatz, we obtain a one-dimensional Schrödinger equation for $\psi(y)$:

$$\mathcal{H}_{k_x}(y) \psi(y) \doteq \mathcal{E}_{k_x}(y) \psi(y), \quad (4.31)$$

where the Hamiltonian $\mathcal{H}(k_x, y)$ is obtained from the massless Dirac Hamiltonian $\mathcal{H}_0(\mathbf{k})$ by replacing $k_y \rightarrow -i\hbar\partial_y$ due to the explicit dependence of the vector potential on the y coordinate:

$$\mathcal{H}_{k_x}(y) = \hbar v_F \left[\left(k_x - \frac{eB_\perp}{\hbar c} y \right) \sigma_x + i\partial_y \sigma_y \right] \quad (4.32a)$$

$$= \hbar v_F \begin{bmatrix} 0 & k_x - \frac{eB_\perp}{\hbar c} y + \partial_y \\ k_x - \frac{eB_\perp}{\hbar c} y - \partial_y & 0 \end{bmatrix}. \quad (4.32b)$$

Similar to the quantum Hall effect in a 2D electron gas (cf. chapter 2) let us introduce the **magnetic length** l_B defined by

$$l_B \equiv \sqrt{\frac{\hbar}{e|B_\perp|}}. \quad (4.33)$$

As a consequence, the Dirac Hamiltonian $\mathcal{H}_{k_x}(y)$ can be rewritten as

$$\mathcal{H}_{k_x}(y) = \frac{\hbar v_F}{l_B} \left[\left(l_B k_x - \text{sgn}(B_\perp) \frac{y}{l_B} \right) \sigma_x + i l_B \partial_y \sigma_y \right] \quad (4.34a)$$

$$= \frac{\hbar v_F}{l_B} \begin{bmatrix} 0 & l_B k_x - \text{sgn}(B_\perp) \frac{y}{l_B} + l_B \partial_y \\ l_B k_x - \text{sgn}(B_\perp) \frac{y}{l_B} - l_B \partial_y & 0 \end{bmatrix}. \quad (4.34b)$$

To solve this Hamiltonian it is convenient to introduce a new dimensionless coordinate q ,

$$q \equiv l_B k_x - \text{sgn}(B_\perp) \frac{y}{l_B}, \quad \partial_q \equiv -\text{sgn}(B_\perp) l_B \partial_y, \quad (4.35)$$

and to rewrite the Dirac Hamiltonian $\mathcal{H}(k_x, y)$ in terms of this new coordinate:

$$\mathcal{H}_{k_x}(q) = \frac{\hbar v_F}{l_B} (q \sigma_x - i \text{sgn}(B_\perp) \partial_q \sigma_y) = \frac{\hbar v_F}{l_B} \begin{bmatrix} 0 & q - \text{sgn}(B_\perp) \partial_q \\ q + \text{sgn}(B_\perp) \partial_q & 0 \end{bmatrix}. \quad (4.36)$$

Now let us introduce a pair of raising and lowering operators $\mathcal{O}, \mathcal{O}^\dagger$ similar to the quantum mechanical treatment of the harmonic oscillator. In the case of Dirac fermions, those raising and lowering operators are given by

$$\mathcal{O}^\dagger \equiv \frac{q - \partial_q}{\sqrt{2}}, \quad \mathcal{O} \equiv \frac{q + \partial_q}{\sqrt{2}}, \quad (4.37)$$

which satisfy the usual commutation relations for bosonic raising and lowering operators:

$$[\mathcal{O}, \mathcal{O}^\dagger] = 1. \quad (4.38)$$

Finally, in analogy to the cyclotron motion of the Dirac fermions within the xy plane let us introduce a cyclotron frequency ω_c for Dirac fermions given by:

$$\omega_c \equiv \sqrt{2} \frac{\hbar v_F}{l_B} = \sqrt{\frac{2v_F^2 e |B_\perp|}{\hbar}}. \quad (4.39)$$

As a result, the Hamiltonian for Dirac fermions moving in the xy plane subjected to a perpendicular magnetic field $\mathbf{B} = B_\perp \mathbf{e}_z$ can be reformulated in terms of raising and lowering operators $\mathcal{O}, \mathcal{O}^\dagger$ as

$$\mathcal{H}_{k_x}^- = \hbar \omega_c \begin{bmatrix} 0 & \mathcal{O} \\ \mathcal{O}^\dagger & 0 \end{bmatrix} \quad \text{and} \quad \mathcal{H}_{k_x}^+ = \hbar \omega_c \begin{bmatrix} 0 & \mathcal{O}^\dagger \\ \mathcal{O} & 0 \end{bmatrix}, \quad (4.40)$$

where the superscript “ \pm ” indicates the direction of the perpendicular magnetic field given by $\text{sgn}(B_\perp)$. To compute the eigenenergies or Landau levels for Dirac fermions, let us calculate the square of the above Hamiltonians $(\mathcal{H}_{k_x}^\pm)^2$:

$$(\mathcal{H}_{k_x}^-)^2 = (\hbar \omega_c)^2 \begin{bmatrix} \mathcal{O} \mathcal{O}^\dagger & 0 \\ 0 & \mathcal{O}^\dagger \mathcal{O} \end{bmatrix} = (\hbar \omega_c)^2 \begin{bmatrix} \hat{n} + 1 & 0 \\ 0 & \hat{n} \end{bmatrix}, \quad (4.41)$$

where we have used the fact that $\mathcal{O}, \mathcal{O}^\dagger$ obey bosonic commutation relations [cf. Eq. (4.38)] to express the Hamiltonian in terms of the number operator $\hat{n} = \mathcal{O}^\dagger \mathcal{O}$. Similarly, for $\mathcal{H}_{k_x}^+$ we obtain:

$$(\mathcal{H}_{k_x}^+)^2 = (\hbar\omega_c)^2 \begin{bmatrix} \mathcal{O}^\dagger \mathcal{O} & 0 \\ 0 & \mathcal{O} \mathcal{O}^\dagger \end{bmatrix} = (\hbar\omega_c)^2 \begin{bmatrix} \hat{n} & 0 \\ 0 & \hat{n} + 1 \end{bmatrix}. \quad (4.42)$$

Hence, we can easily construct the eigenfunctions of $(\mathcal{H}_{k_x}^\pm)^2$ using the eigenfunctions $|n\rangle$ of the number operator \hat{n} . For the perpendicular magnetic field oriented along the $+z$ direction, *i.e.*, for $\text{sgn}(B_\perp) = 1$, we obtain a set of eigenfunctions of $(\mathcal{H}_{k_x}^+)^2$ given by

$$\psi_{n=0}^+ \propto \begin{bmatrix} |0\rangle \\ 0 \end{bmatrix} \quad \text{and} \quad \psi_{n>0}^+ \propto \begin{bmatrix} |n\rangle \\ |n-1\rangle \end{bmatrix}, \quad (4.43)$$

while for a magnetic field oriented along the $-z$ direction with $\text{sgn}(B_\perp) = -1$ the set of eigenfunctions of the Dirac Hamiltonian $(\mathcal{H}_{k_x}^-)^2$ given by

$$\psi_{n=0}^- \propto \begin{bmatrix} 0 \\ |0\rangle \end{bmatrix} \quad \text{and} \quad \psi_{n>0}^- \propto \begin{bmatrix} |n-1\rangle \\ |n\rangle \end{bmatrix}. \quad (4.44)$$

In both cases, however, the eigenenergies or Landau levels of Dirac fermions moving in a perpendicular magnetic field are independent of the longitudinal momentum k_x . As $(\mathcal{H}_{k_x}^\pm)^2 \psi_n^\pm = n(\hbar\omega_c)^2 \psi_n^\pm$, the eigenenergies of the original Dirac Hamiltonian \mathcal{H}_0 are given by

$$\mathcal{E}_n = \hbar\omega_c \text{sgn}(n) \sqrt{|n|} = \text{sgn}(n) \sqrt{2v_F^2 e \hbar |B_\perp| |n|}. \quad (4.45)$$

In contrast to the quantum Hall effect of a 2D electron gas (cf. chapter 2), the distance between the Landau levels is not constant, but decreases with increasing Landau level index n as $1/\sqrt{|n|}$. Moreover, note that there exists a half-filled Landau level at zero energy, *i.e.*, $\mathcal{E}_{n=0} = 0$, due to particle-hole symmetry.

Now let us consider the effect of a mass term on the Landau level spectrum. In the presence of a Dirac mass term such as $m_{\text{surf}} \sigma_z$, the ground state degeneracy is lifted, because the ground state wave functions ψ_0^\pm acquire a finite energy with opposite sign for the different orientations of the magnetic field:

$$m_{\text{surf}} \sigma_z \psi_0^\pm = \pm m_{\text{surf}} \psi_0^\pm = \text{sgn}(B_\perp) m_{\text{surf}} \psi_0^\pm. \quad (4.46)$$

On the other hand, the higher Landau levels obtain an additional energy correction $\propto m_{\text{surf}}$ since $\sigma_z^2 = \mathbb{1}$:

$$\mathcal{E}_n = \text{sgn}(n) \sqrt{2e\hbar |B_\perp| |n| + m_{\text{surf}}^2} \quad \text{for } n \neq 0. \quad (4.47)$$

To conclude, the **Landau level sequence** for massive Dirac fermions is given by

$$\mathcal{E}_n = \begin{cases} \text{sgn}(B_\perp) m_{\text{surf}} & \text{for } n = 0 \\ \text{sgn}(n) \sqrt{2v_F^2 e \hbar |B_\perp| |n| + m_{\text{surf}}^2} & \text{for } n \neq 0 \end{cases} \quad (4.48)$$

Note, however, that the correction of the Landau levels due to the Dirac mass term is only relevant for the zeroth Landau level, because usually the orbital Landau level splitting $\hbar\omega_c$ is much bigger than the Zeeman spin-splitting, so that we can neglect this small energy correction for all but the zeroth Landau level:

$$m_{\text{surf}} \sim \mu_B \|\mathbf{B}_Z\|_2 \ll \hbar\omega_c \sim \hbar v_F \sqrt{e |B_\perp|}. \quad (4.49)$$

In that sense, the Landau levels for massive Dirac fermions are similar to those of massless Dirac fermions [cf. Eq. (4.45)], except that the zeroth Landau level is not at zero energy anymore.

The above results are valid for any planar surface of the strong topological insulator with surface normal vector \hat{n} and normal magnetic field component $B_{\perp} = \hat{n} \cdot \mathbf{B}$. Since the Landau levels (4.48) do not cross as a function of the applied magnetic field B_{\perp} or the surface Dirac mass term m_{surf} , they are independent of quantitative details with respect to m_{surf} . We may then characterize each surface by a topological invariant or Hall quantum number ν defined by

$$\nu \equiv \begin{cases} n + 1 & \text{for } B_{\perp} > 0 \\ -n & \text{for } B_{\perp} < 0 \end{cases} \quad (4.50)$$

where we have assumed that the chemical potential μ_{surf} on that surface lies in between adjacent Landau levels, $\mathcal{E}_n < \mu_{\text{surf}} < \mathcal{E}_{n+1}$. Experimentally, the surface chemical potential μ_{surf} can, in principle, be controlled via gate voltages. Fig. 4.4 shows the Landau level sequence (4.48) and the corresponding Hall quantum number ν calculated from Eq. (4.50) as function of the surface Dirac mass term m_{surf} for different magnetic field orientations ($B_{\perp} > 0$ and $B_{\perp} < 0$). Note that the splitting of the zeroth Landau level due to m_{surf} may lead to different quantum numbers $\nu = 0$ or $\nu = 1$ depending on the relative sign of the mass term, if the chemical potential μ is tuned appropriately.

For a single surface, the quantitative details of the Dirac mass term and the magnetic field orientation do not matter as long as they change adiabatically in the sense that the quantum number ν , which characterizes the surface, does not change. If, however, we consider a finite-size sample with planar surfaces, there exist a number of edges, where two surfaces intersect. In particular, if two surfaces with different quantum numbers ν intersect, there appear edge states due to the bulk-boundary correspondence, as we discuss below. Furthermore, in a finite sample the edge states are connected, leading to edge channel networks, and the dependence of the Landau level sequence on the Dirac mass term m_{surf} leads to interesting physics when the orientation of the magnetic field is changed, as will be discussed later as well.

4.1.5 The appearance of edge states

To illustrate the emergence of chiral edge states in the presence of an external magnetic field, let us consider an edge, where two surfaces intersect. Assuming that both surfaces have a finite normal magnetic field component B_{\perp} , the electronic states at those surfaces form Landau levels, and we can characterize the surfaces by Hall quantum numbers ν_1 and ν_2 defined by Eq. (4.50), as previously discussed.

Furthermore, the edge can be considered as a 1D “domain wall” between 2D surfaces of a finite-size 3D sample of a topological insulator, across which some parameter varies spatially. In particular, this spatial phase transition may result in gapless excitations which form in the vicinity of the edge of a topological insulator. This concept known as **bulk-boundary correspondence** relates the topological properties of the surfaces, *i.e.*, the quantum number ν , to the number of gapless edge modes. Other examples of the bulk-boundary correspondence include the edge states in the Haldane model discussed in chapter 2 or the topological excitations at the domain walls in the Su-Schrieffer-Heeger model (cf. appendix C).

In the context of an edge between two planar surfaces of the minimal model in the presence of a magnetic field, there exist $\Delta\nu = \nu_1 - \nu_2$ edge states if the quantum numbers ν_1 and ν_2 differ, there exist quantum Hall edge states, and their direction is given by the relative magnitude of ν_1 and ν_2 . Using Eqs. (4.48) and (4.50) we can therefore calculate the Hall quantum numbers ν for a quasi-1D beam-shaped sample of a strong topological insulator for a finite-size

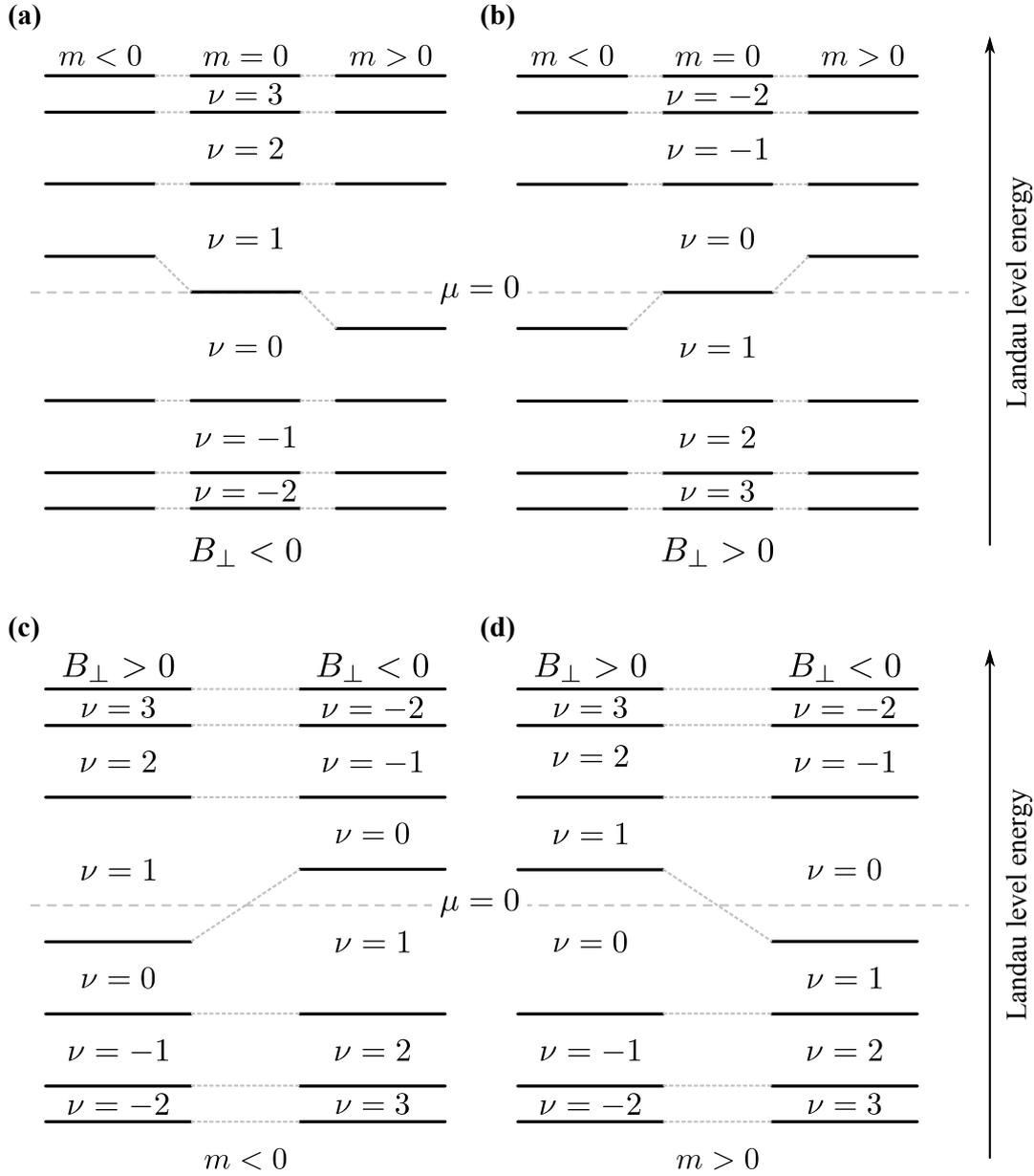


Figure 4.4. Schematic picture of the Landau level sequence (4.48) as function of the surface Dirac mass term $m = m_{\text{surf}}$ and the normal magnetic field B_{\perp} . The dashed lines are guides to the eye, connecting Landau levels belonging together. Panels (a) and (b) show the same Landau level sequence for opposite orientations of the normal magnetic field component ($B_{\perp} > 0$ and $B_{\perp} < 0$). In the absence of a Dirac mass term, the $n = 0$ Landau level has zero energy and is located in the center of the sequence. The corresponding Hall quantum number ν has been calculated from Eq. (4.50). If the chemical potential μ is close to the zeroth Landau level, one can obtain different quantum numbers for a surface, depending on the relative sign of the mass term m . Panels (c) and (d) show the Landau level sequences for opposite top and bottom surfaces with normal vectors $\hat{n}_t = -\hat{n}_b$. As a consequence, the normal magnetic fields at the top and bottom surfaces, $B_{\perp,t}$ and $B_{\perp,b}$, have opposite sign, and the sequences of Hall quantum numbers run in opposite direction.

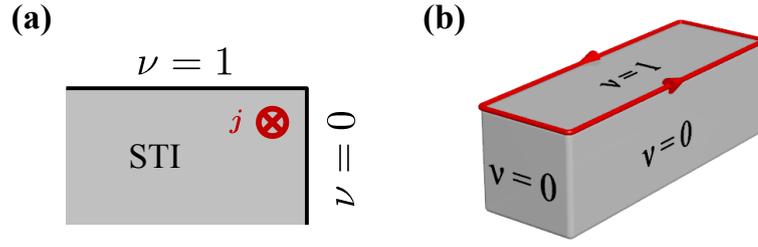


Figure 4.5. (a) Schematic view of an edge state in the direction of the current. The planar surfaces of the topological insulator are characterized by two quantum numbers ν_1 and ν_2 as defined by Eq. (4.50) in the main text. The number and orientation of edge states is then given by $\Delta\nu = \nu_1 - \nu_2$. (b) Illustration of a current-carrying edge channel running along the edges of a finite 3D topological insulator in a magnetic field.

sample as shown in Fig. 4.5 (a). Then, the differences $\Delta\nu = \nu_1 - \nu_2$ determine the position and number of the edge channels, where ν_1 and ν_2 characterize two adjacent surfaces which intersect. Moreover, concerning a finite 3D sample of a topological insulator, those edge states connect and form closed current loops, as illustrated in Fig. 4.5 (b).

To develop a better understanding of the edge states and their properties, let us consider different limits of model parameters. As an example, Fig. 4.6 shows the appearance of Hall currents propagating along the one-dimensional interfaces along the x and $-x$ direction for a tilted magnetic field \mathbf{B} . For $\mu = 0$, the relevant Landau levels close to the Fermi energy are the $n = 0$ and $n = \pm 1$ Landau levels, respectively. The position of the induced edge currents, however, depends on the sign of the surface Dirac mass m_{surf} . By tuning the surface chemical potential such that $m_{\text{surf}} \approx 0$, the Zeeman interaction and inversion asymmetry are small perturbations, so that the edge states are dominated by orbital magnetic field effects. As a consequence, the edge states are located in a plane normal to the magnetic field direction, as shown in Fig. 4.6 (a), which is similar to the usual quantum Hall effect of a 2D electron gas. On the other hand, when $m_{\text{surf}} \approx (g_- B_Z^x, g_- B_Z^y, -g_+ B_Z^z)^T \cdot \hat{\mathbf{n}}$ the Dirac mass term explicitly depends on the surface orientation relative to the magnetic field. In the beam-shaped sample of Fig. 4.6, this leads to a sign change of the Dirac masses on opposite surfaces, effectively interchanging the Hall quantum numbers of the top and bottom surfaces. The resulting counter-propagating edge currents are thus located at opposite edges in the direction of the applied magnetic field \mathbf{B} [cf. Fig. 4.6 (b)], which is similar to the Zeeman spin-splitting of spin-up and spin-down states.

4.1.6 Edge channel networks

As mentioned in the previous section, since all surfaces of a finite 3D topological insulator are connected, the chiral edge states are also connected to each other, thus forming **networks** of edge channels. Due to the dependence of the Dirac mass term on the relative orientation of the magnetic field and the planar surfaces of the sample, this can lead to interesting networks of edge states when the direction of the magnetic field is changed.

To be specific, we have calculated the quantum numbers ν [cf. Eq. (4.50)] and the number of edge channels $\Delta\nu$ for the labeled edges in the finite-size geometry shown in Fig. 4.7 (d). Here, the magnetic field was chosen such that it can be parametrized as $\mathbf{B} = B_0(1, \cos \phi, \sin \phi)^T$, where ϕ denotes the rotation angle with respect to the y axis [cf. Fig. 4.7 (a)]. This particular realization of a magnetic field allows a wide variety of edge channel networks.

For example, note that for $\phi = n\pi$ with integer n the magnetic field is perpendicular to the z direction, *i.e.*, the top and bottom surfaces with normal vectors $\pm \mathbf{e}_z$ are not gapped by the

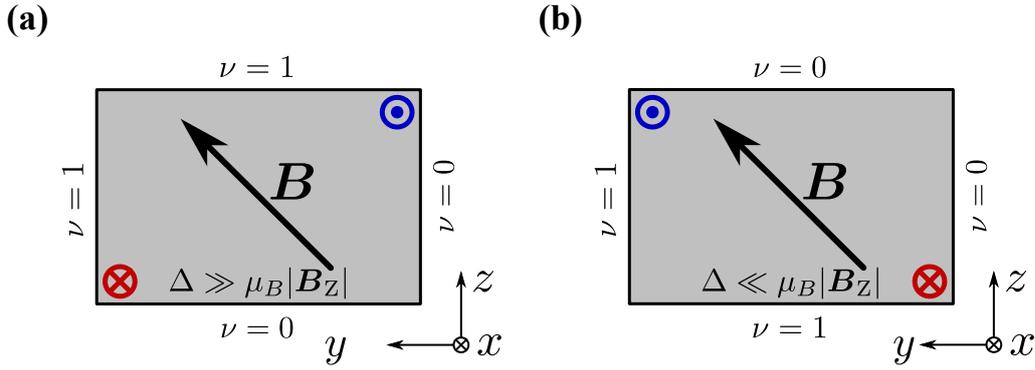


Figure 4.6. Schematic cross section of a beam-shaped sample of a strong topological insulator with infinite extension in x direction, but finite height and width in y and z directions. B denotes the applied magnetic field responsible for the Landau level spectrum of the surface Dirac fermions. Calculating the Hall quantum numbers ν for each surface using Eqs. (4.48) and (4.50), surfaces with different quantum numbers ν_1 and ν_2 induce Hall currents along their shared edges, and their number and orientation is given by $\Delta\nu = \nu_1 - \nu_2$. **(a)** Tuning the surface chemical potential such that $m_{\text{surf}} \approx 0$, the Zeeman interaction and inversion asymmetry are small perturbations, and the positions of the Hall currents are dominated by the orbital field, *i.e.*, they are localized in a plane perpendicular to the magnetic field. **(b)** If the Zeeman interaction dominates, *i.e.*, $m_{\text{surf}} \approx (g_- B_Z^x, g_- B_Z^y, -g_+ B_Z^z)^T \cdot \hat{n}$, then the mass terms on opposite surfaces of the sample have different sign, leading to different Hall quantum numbers ν at the top and bottom surfaces. In contrast to the orbitally dominated edge states, the induced Hall currents are located at the edges along the direction of the magnetic field, showing a Zeeman-like character.

applied magnetic field. As a consequence, in that limit the cyclotron frequency for the surface Dirac fermions vanishes, $\omega_c \rightarrow 0$, and the Hall quantum numbers ν of the top and bottom surfaces diverge as $\phi \rightarrow n\pi$. Similarly, for $\phi = (n + \frac{1}{2})\pi$ with integer n , the magnetic field is oriented parallel to the left and right side surfaces with normal vectors in $\pm y$ direction.

For intermediate rotation angles ϕ we observe, however, much more interesting edge channel configurations. We have calculated the number of edge channels as function of the rotation angle ϕ of the magnetic field using Eq. (4.50), with the results for the different edges plotted in Fig. 4.7 **(a)**. For instance, for $\phi = \pi/4$ the magnetic field points in the $(\sqrt{2}, 1, 1)^T$ direction, and we observe a closed edge channel in a plane almost perpendicular to this direction which is dominated by the orbital effects of the magnetic field, as shown in Fig. 4.7 **(b)**. This is similar to the orbitally dominated current-carrying edge states, as discussed previously and shown in Fig. 4.6 **(a)**. If we rotate the magnetic field further towards the x axis, *i.e.*, for $\phi = \pi/15$, the cyclotron frequencies of the top and bottom surface decrease which leads to a “contraction” of the corresponding Landau level sequence. Hence, for fixed surface chemical potential μ_{surf} the Hall quantum numbers of the top and bottom surface increase (decrease), and this leads to the appearance of additional closed edge channels around the top and bottom surfaces, as shown in Fig. 4.7 **(c)**.

Moreover, in principle, one can control the chemical potential of each surface of the topological insulator sample by applying a gate voltage through side contacts. By changing the chemical potential of the surfaces, we can construct similar sequences of quantum Hall transitions on the surfaces. The different patterns of those **edge channel networks** obtained upon changing the strength of the magnetic field, its orientation, or the side gate voltages, may thus serve as a fingerprint of a topological insulator and the related Dirac-like surface states, in contrast to the more conventional chiral edge channels observed in the usual quantum Hall samples. It also opens up the opportunity to study quantum Hall physics by separately contacting

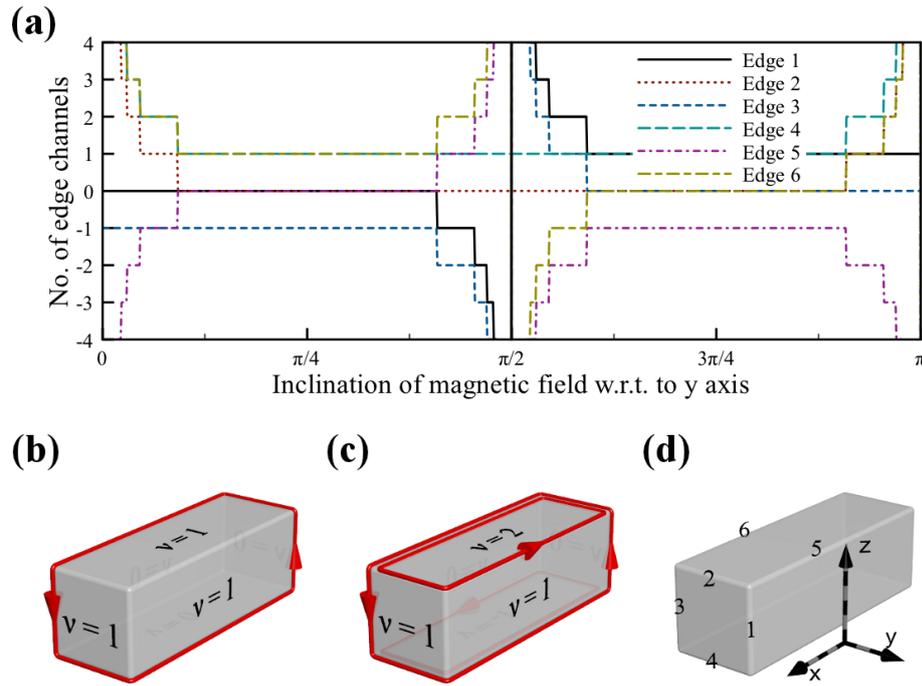


Figure 4.7. (a) The number of edge channels along the edges 1–6 as defined in panel (d) for a beam-shaped sample of a strong topological insulator. The Hall quantum numbers ν and their differences corresponding to the edges have been calculated from Eq. (4.50) as a function of the inclination angle ϕ for the chemical potential $\mu/t = 0.2$ using the model parameters $t = 1$, $\Delta/t = 0$, $\mathbf{B} = \frac{0.001}{\sqrt{2}}(1, \cos \phi, \sin \phi)^T$. The number of edge channels on the six other edges are obtained by inversion symmetry since $\Delta/t = 0$, and the sign indicates whether the current runs parallel or anti-parallel to one of the Cartesian coordinate axes. At $\phi = 0$ and $\phi = \frac{\pi}{2}$ the magnetic field is perpendicular to either the y or z direction. For the two inclination angles $\phi = \frac{\pi}{4}$ and $\phi = \frac{\pi}{15}$ the edge channel configurations are shown in panels (b) and (c), respectively. On each face of the sample, the Hall quantum number ν is written which characterizes the surface, and the number of edge channels is obtained as the difference $\nu_1 - \nu_2$ of the surfaces intersecting at the edges. Panel (d) indicates the numbering of edges and the coordinate system. We have published parts of this figure in Ref. 19.

the edge channels which, for example, might allow for the study of heat transport in topological insulators¹⁹.

4.1.7 Conductance networks

For a given edge channel network we can also calculate the corresponding **conductance network** due to the fact that each edge channel carries a current defined by the relation $I/V = G_0$. Such a network of conductances could, in principle, be studied by applying contacts at the individual edges of the sample and measuring the voltage drop between the different contacts.

As an explicit example, let us consider the edge channel network shown in Fig. 4.7 (c) and project the network onto a two-dimensional plane as shown in Fig. 4.8, where the Hall quantum numbers ν as calculated from Eq. (4.50) characterize the individual surfaces of the beam-shaped sample, and $\nu = 0$ on the outside of the sample. The difference $\Delta\nu = \nu_1 - \nu_2$ between surfaces which intersect then determines the number and direction of the edge channel states as described above.

Using the fact that the total current entering and leaving a contact is equal (Kirchhoff's law) and that the electrochemical potential remains constant when channels split, we can calculate

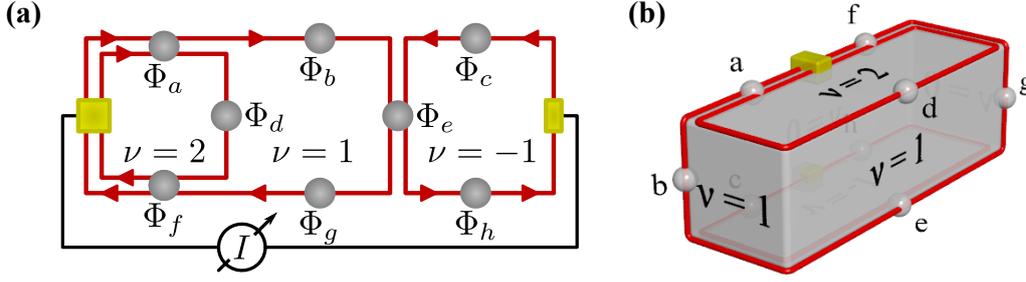


Figure 4.8. Illustration of the conductance network corresponding to the non-trivial edge channel network shown in Fig. 4.7 (c) which is obtained by projecting the edge channels onto a two-dimensional plane. The Hall quantum numbers ν characterize the surfaces, with the outside of the network corresponding to $\nu = 0$. By applying a bias voltage at the rectangular (golden) contacts, a current I is passed through the sample which is carried by the chiral edge channels. The electrochemical potentials Φ at the individual voltage contacts $a-h$ (indicated by gray circles) allow, in principle, to study the conductance network. We have published a similar figure in Ref. 19.

the electrochemical potentials Φ_i at the edge contacts $a-h$ as follows. Injecting a current I at the left contact indicated by a golden square, the contacts Φ_a , Φ_b , and Φ_d all lie on the same electrochemical potential Φ . With respect to the potential Φ_a at contact a a voltage drop I/G_0 occurs at contacts e and g , because there the number of edge channels has reduced from two to one. A similar reasoning for the other contacts then shows that the conductance network shown in Fig. 4.8 is characterized by

$$\begin{aligned} & (\Phi_a, \Phi_b, \Phi_c, \Phi_d, \Phi_e, \Phi_f, \Phi_g, \Phi_h)^T \\ & = (\Phi, \Phi, \Phi - 2I/G_0, \Phi, \Phi - I/G_0, \Phi - \frac{1}{2} I/G_0, \Phi - I/G_0, \Phi - I/G_0)^T, \end{aligned} \quad (4.51)$$

where $G_0 \equiv e^2/h$ is the elementary conductance quantum. Defining the skew-symmetric matrix of cross-conductances \mathcal{G} by

$$\mathcal{G}_{\alpha\beta} \equiv \frac{I}{\Phi_\beta - \Phi_\alpha}, \quad (4.52)$$

a straightforward calculation then shows that

$$\mathcal{G} = \frac{e^2}{h} \begin{bmatrix} \infty & \infty & -\frac{1}{2} & \infty & -1 & -2 & -1 & -1 \\ \infty & \infty & -\frac{1}{2} & \infty & -1 & -2 & -1 & -1 \\ \frac{1}{2} & \frac{1}{2} & \infty & \frac{1}{2} & 1 & \frac{2}{3} & 1 & 1 \\ \infty & \infty & -\frac{1}{2} & \infty & -1 & -2 & -1 & -1 \\ 1 & 1 & -1 & 1 & \infty & 2 & \infty & \infty \\ 2 & 2 & -\frac{2}{3} & 2 & -2 & \infty & -2 & -2 \\ 1 & 1 & -1 & 1 & \infty & 2 & \infty & \infty \\ 1 & 1 & -1 & 1 & \infty & 2 & \infty & \infty \end{bmatrix}. \quad (4.53)$$

Note that the infinite matrix elements of the cross-conductance tensor stem from the fact that we assume a vanishing bulk conductance. A close inspection of the cross-conductance tensor furthermore reveals that depending on the experimental setup one cannot only measure integer conductances such as $\mathcal{G}_{gb} = G_0$ or $\mathcal{G}_{fa} = 2G_0$, but also rational values, for instance $\mathcal{G}_{cf} = \frac{2}{3}G_0$ (Ref. 19). As a consequence, a single experiment probing the edge channel network can be misleading in the sense that the experimental results depend very much on the positions of the voltage probes relative to the contacts, where the bias voltage driving the current is applied, on the orientation of the magnetic field, and on the surface chemical potentials.

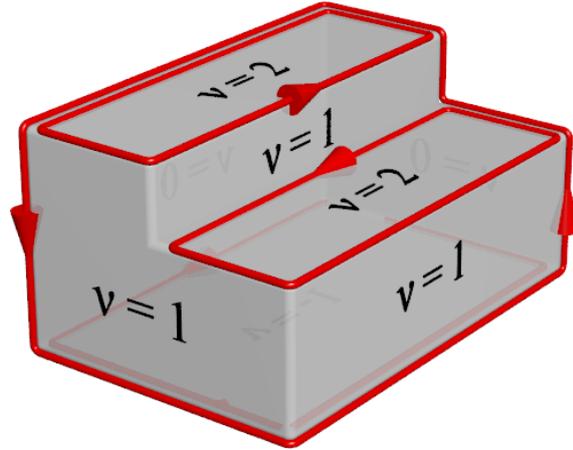


Figure 4.9. Schematic illustration of a topological insulator with a sharp step edge. Due to the different Hall quantum numbers ν on the surfaces, the corresponding chiral channels propagate in opposite directions at the step edge. This allows, in principle, to study the interactions and scattering processes between counter-propagating edge channels as function of the step height.

Nevertheless, the particular dependence of the full cross-conductance tensor \mathcal{G} as function of the rotating magnetic field allows, in principle, for an experimental identification of topological insulators.

Furthermore, due to the quantization of the charge transport within the edge channels, the power P which is dissipated at the points, where two edge channels meet, is also quantized. In other words, in such a conductance network one could, in principle, observe a quantized Ohmic conductance I^2/P . For example, according to the above discussion of the electrochemical potential differences, the dissipated power close to the contact e takes the quantized value $2G_0$, whereas close to contact f the dissipated power P takes on the value $G_0/2$. Such a quantized heat source might also have interesting applications for the study of heat transport in topological insulators.

4.1.8 Counter-propagating edge channels at step edges

Another interesting aspect of such edge channel networks are the interactions and scattering processes of chiral edge states located at step edges on the surface of a strong topological insulator, as shown in Fig. 4.9. Since the edge states are chiral, *i.e.*, the electrons move in only direction at one edge and in opposite direction at the other edge, there is no backscattering at the same edge. However, the conductance is quite sensitive to tunneling between the counter-propagating edge channels. Note that any charge that tunnels between the two edge states is effectively backscattered which leads, in principle, to a reduction of the conductance. Moreover, since the edge channel networks depend sensitively on both the surface chemical potential and the orientation of the external magnetic field with respect to the sample, this setup allows basically to study the interaction effects between different numbers of edge channels.

In section 4.2.2 below, we discuss the numerical results for the electronic band structure and the localized edge channels obtained for different heights of the step edge using the tight-binding approach. We also briefly discuss the Coulomb interaction between the two chiral edge channels at the step edge.

4.2 Numerical results for the quantum Hall effect in the minimal model

In the previous section, we have obtained analytical results for the Landau levels on the surfaces of a strong topological insulator subjected to an external magnetic field. Notably, we have shown that the edge channels in a finite-size geometry show a particular dependence on the relative orientation of the magnetic field with respect to the sample due to a non-universal Dirac mass term in the surface Dirac Hamiltonian. To illustrate (i) the locations of the chiral edge channels along the edges of a 3D sample subjected to an external magnetic field, (ii) the appearance of the edge channel networks upon rotating the magnetic field, and (iii) the counter-propagating edge channels at the step edges, we have calculated the properties of an infinite beam-shaped sample and an infinite slab using exact diagonalization techniques. In this section, we discuss the numerical results and compare them with the analytical ones obtained in the previous section. Here, it turns out that we obtain a very good agreement between the analytical and numerical results up to a few percent in magnitude.

4.2.1 Numerical results for the chiral edge states

First, let us consider the electronic band structure of a finite-size geometry with rectangular cross section and dimensions $N_x \times N_y \times N_z = \infty \times 50 \times 40$. Although the orbital magnetic field breaks translation symmetries, in the appropriate Landau gauge the momentum along the x direction is conserved, and we may consider a two-dimensional tight-binding problem in the yz plane parametrized by the momentum k_x [cf. Eq. (4.12)]. The band structure of such a tight-binding Hamiltonian can be calculated by, for example, exact diagonalization. Fig. 4.10 shows the resulting band structure as a function of the momentum k_x in different limits which allows to study the effect of the individual terms entering the two-dimensional tight-binding Hamiltonian (4.12).

Let us start with the band structure of the clean model, *i.e.*, in the absence of any symmetry-breaking terms or an external magnetic field, as shown in Fig. 4.10 (a) for $m/t = 2$, $\Delta/t = 0$, $N_\Phi^y = N_\Phi^z = 0$, and $\mathbf{B}_Z = 0$. In that limit, the cyclotron frequency vanishes, and each surface exhibits Dirac-like surface states which can be easily observed in panel (a). Note that the tiny band gap stems from finite-size effects and vanishes as the cross section $N_y \times N_z \rightarrow \infty$. A Dirac mass term such as the symmetry-breaking term Δ , which is constant and equally strong on all surfaces of the sample, leads to a massive Dirac spectrum, *i.e.*, a band gap of the order of 2Δ [compare Fig. 4.10 (a, b)].

By applying an external magnetic field, Landau levels emerge, and the spectrum of the Dirac fermions changes, as discussed in the previous section. If the magnetic field is oriented such that two of the side surfaces remain gapless, the band structure exhibits both Dirac-like states and flat Landau levels, as shown in 4.10 (c). To observe a well-defined quantum Hall signature, the magnetic field has to be oriented in such a way that there is a finite normal magnetic field component on all side surfaces. In that case, all electronic surface states form Landau levels which are represented by the flat parts in the electronic band structure in Fig. 4.10 (d), while the dispersive parts correspond to the 1D chiral edge channels. To obtain the positions of the chiral edge channels, we have calculated the corresponding probability distributions in the absence of Zeeman interactions, as shown in Fig. 4.10 (e, f), where the filled right (left) triangles in red (blue) color indicate particles moving in positive (negative) x direction, respectively. As expected from the analytical approach, the current-carrying edge channels are located at the edges within a plane normal to the magnetic field direction, and their relative position depends on the chemical potential above or below zero energy [cf. Fig. 4.10 (e, f)].

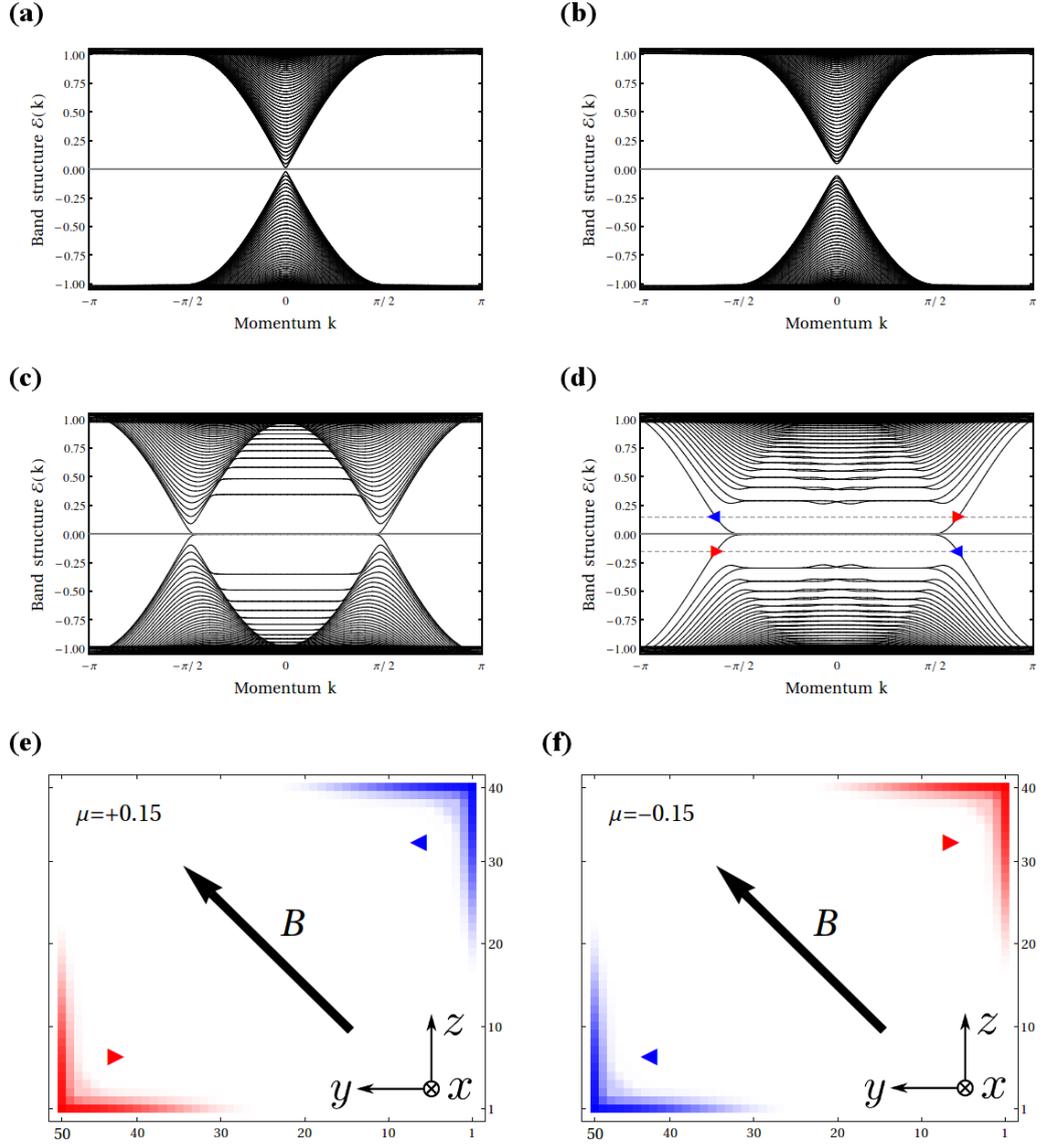


Figure 4.10. Plots of the electronic band structure of a beam-shaped strong topological insulator with dimensions $N_x \times N_y \times N_z = \infty \times 50 \times 40$. The parameters of the model have been chosen such that it realizes a strong topological insulator in the absence of magnetic fields: $t = 1$, $m/t = 2$, $N_y = 40$, $N_z = 50$, $g_{\pm} = \frac{1}{2}$, $\mathbf{B}_Z = 0$. **(a)** In the absence of an orbital magnetic field ($N_{\Phi}^y = N_{\Phi}^z = 0$) and inversion asymmetry ($\Delta/t = 0$) we observe Dirac-like surface states on all four side surfaces. **(b)** In the presence of a Dirac mass term $\Delta \sigma_z$ with $\Delta = 0.05$, those Dirac surface states become massive, opening up a band gap. **(c)** By applying an orbital magnetic field, the Dirac fermions are re-organized into Landau levels. If the magnetic field is oriented parallel to one of the side surfaces ($\Delta/t = 0$, $N_{\Phi}^y = 0.01$, $N_{\Phi}^z = 0$), this surface remains gapless, leading to the formation of surface states resembling a Dirac cone. **(d)** Upon rotating the orbital magnetic field such that all surfaces exhibit a finite normal magnetic field component ($\Delta/t = 0$, $N_{\Phi}^y = N_{\Phi}^z = 0.007$), the electronic states of all side surfaces form Landau level sequences. The flat parts represent the Landau levels, while the dispersive parts correspond to the one-dimensional edge channels. **(e, f)** Logarithmic plot of the probability distribution $|\psi(\mathbf{r})|^2$ as function of the two-dimensional coordinate \mathbf{r} in a cross section of the beam-shaped sample for different values of the chemical potential μ/t . Particles moving in positive (negative) x direction are indicated by a filled right (left) triangle in red (blue) color. For positive (negative) μ the edge channels show the expected orbital motion of electrons (holes) with negative (positive) charge.

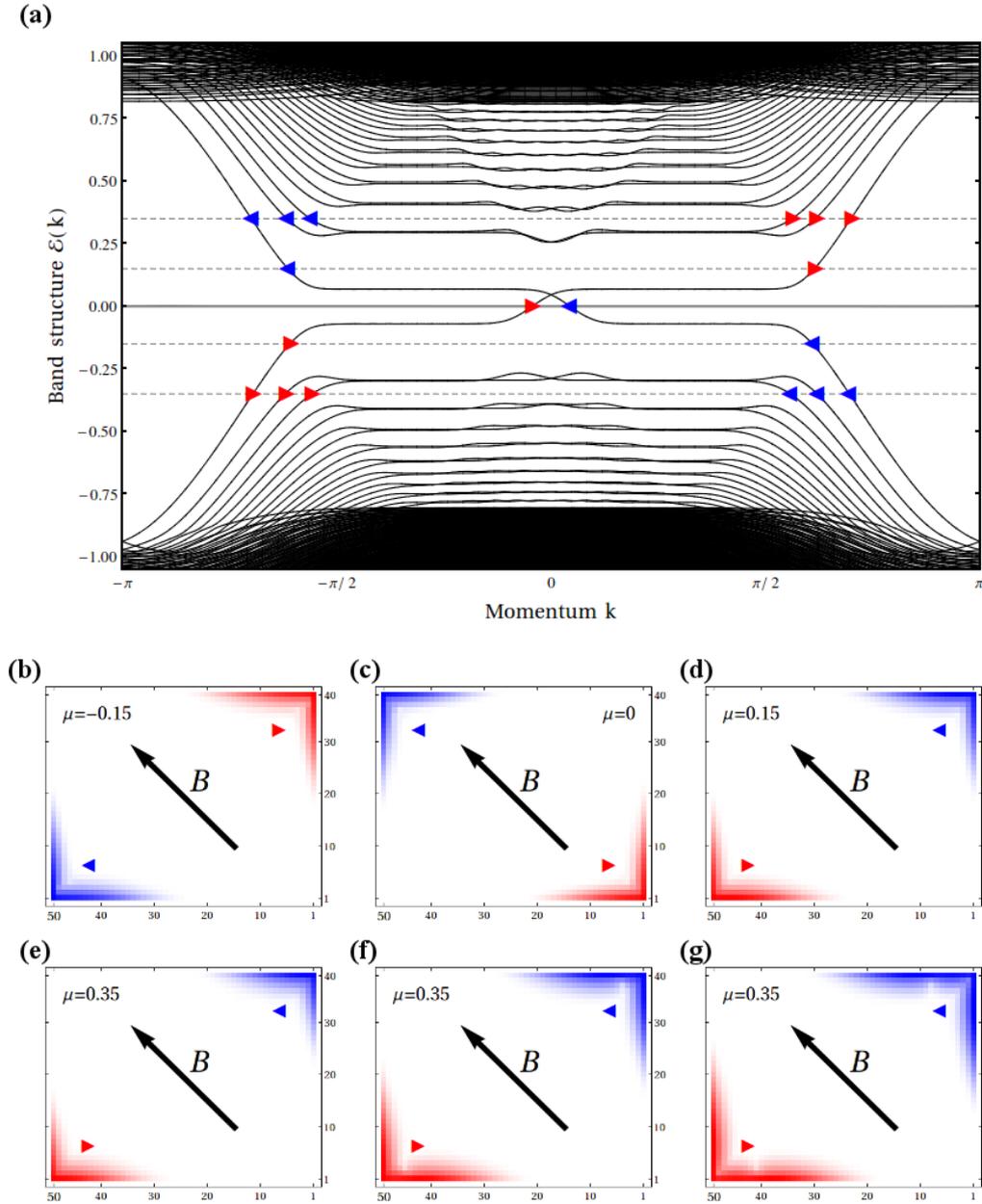


Figure 4.11. (a) Plot of the electronic band structure of a beam-shaped strong topological insulator with dimensions $N_x \times N_y \times N_z = \infty \times 50 \times 40$. The parameters of the model have been chosen such that it realized an inversion-symmetric strong topological insulator in the absence of magnetic fields: $t = 1$, $m/t = 2$, $\Delta/t = 0$, $N_y = 50$, $N_z = 40$, $g_{\pm} = \frac{1}{2}$, $\mathbf{B}_Z = (0, 0.14, 0.14)^T$, $N_{\Phi}^y = N_{\Phi}^z = 0.007$. (b–f) Logarithmic plots of the probability distribution $|\psi(\mathbf{r})|^2$ as function of the coordinate \mathbf{r} in a cross section of the beam-shaped sample for different chemical potentials μ/t . Particles moving in positive (negative) x direction are denoted by a filled right (left) triangle in red (blue) color. (b–d) Sequence of “normal” quantum Hall transitions, where opposite edge channels move simultaneously between edges. For positive (negative) μ the edge channels are dominated by the orbital field and show an orbital motion in a plane perpendicular to the magnetic field, where the direction of the current-carrying edge channels is determined by the negative (positive) charge carriers. The splitting of the zeroth Landau level due to the magnetic Zeeman interaction [cf. Eq. (4.48)] leads to an additional Hall transition which is dominated by the Zeeman interaction, *i.e.*, the edge channels are located in the edges along the direction of the magnetic field \mathbf{B} . (e–g) For larger chemical potential we observe multiple orbitally dominated edge channels located in edges perpendicular to the magnetic field. In the logarithmic plots (from left to right) we observe that the edge states corresponding to higher Landau levels in panel (a) are slightly less localized at the edges. We have published a similar figure in Ref. 19.

However, in a more realistic system the magnetic field also couples to the electronic spin degree of freedom which leads to an effective magnetic Zeeman interaction. As shown previously, the Zeeman interaction in topological insulators is different from the one obtained for Dirac fermions in graphene, because it leads to a mass term which depends on the relative orientation of the magnetic field and the surface normal vector [cf. Eqs. (4.27, 4.48)]. To better understand the role of the Zeeman interaction, let us first consider an inversion-symmetric topological insulator with $\Delta/t = 0$. In that case, the degeneracy of the zeroth Landau level at zero energy is lifted by the Dirac mass term $m_{\text{surf}} \equiv (g_- B_Z^x, g_- B_Z^y, -g_+ B_Z^z)^T \cdot \hat{\mathbf{n}}$ induced by the Zeeman interaction. As discussed in the previous section, the edge channels are then located at the edges along the direction of the magnetic field \mathbf{B} , as shown in Fig. 4.6 (b). To check this, we have calculated the electronic band structure and the probability distributions of the eigenstates within the tight-binding approximation in the presence of a magnetic Zeeman field, as shown in Fig. 4.11. Panels 4.11 (b–d) show a sequence of quantum Hall transitions between the lowest Landau levels \mathcal{E}_n with $n \in \{-1, 0, 1\}$. Note that the energy of the Landau levels is in very good agreement with the analytical Landau level sequence (4.48). While the edge channels at the chemical potential $\mu/t = \pm 0.15$ show the expected orbital character, where the electrons move in a plane perpendicular to the applied magnetic field as discussed above, at $\mu/t = 0$ we observe another quantum Hall transition to a state dominated by the Zeeman interaction. This can be easily seen from the orientation of the edge channels along the direction of the magnetic field \mathbf{B} which is in perfect agreement with our analytical arguments. Moreover, the quantum Hall transitions for the individual edge channels occur simultaneously at both edges due to the rotation symmetry of the model for $\Delta/t = 0$, *i.e.*, edge channels located in opposite edges of the sample “jump” between edges at the same time. For larger chemical potentials, *e.g.*, $\mu/t = \pm 0.35$, we also observe three orbitally dominated edge channels corresponding to the three filled Landau levels \mathcal{E}_n with $n \in \{0, \pm 1, \pm 2\}$. More generally, a closer inspection of the electronic band structure shows that always pairs of Landau levels become occupied at a quantum Hall transition, except for the zeroth Landau level, whose degeneracy is eventually lifted by the Dirac mass term. In that sense, we observe an odd-integer quantum Hall effect.

This picture changes when we take a finite inversion asymmetry Δ/t into account which is generated by (i) inversion asymmetry due to the sample growth on a substrate or (ii) the applied magnetic field breaking time-reversal symmetry. In that case, the two gaps induced by the Zeeman interaction and the symmetry-breaking term Δ/t compete with each other and can lead to interesting quantum Hall transitions. For example, the edge channels can switch between an orbitally dominated character and a Zeeman-dominated behavior depending on the relative strength of the two gaps, as was observed in Figs. 4.10 and 4.11. To shed some light into the more complicated interplay between those two competing gaps, we have calculated the electronic band structure of the minimal model in the presence of a finite orbital magnetic field, a finite Zeeman interaction, and a finite inversion-asymmetry Δ/t which is shown in Fig. 4.12 (a). As before, the Landau level energies agree very well with the analytically obtained Landau level sequence. However, if we consider the sequence of quantum Hall transitions close to the zeroth Landau level, we also observe individual edge channels “jumping” from one edge of the sample to another as function of the chemical potential [see Fig. 4.12 (b–d)].

In principle, it is also possible to obtain more complex situations which show both orbitally dominated and Zeeman interaction dominated edge channels, for instance close to the quantum Hall transition populating the first and second Landau levels. This, however, requires fine-tuning of the chemical potential μ and the applied magnetic field \mathbf{B} as the corresponding feature in the band structure is not very pronounced, and thus it is likely not observable in a real-world quantum Hall experiment. Since such sequences of quantum Hall transitions also occur

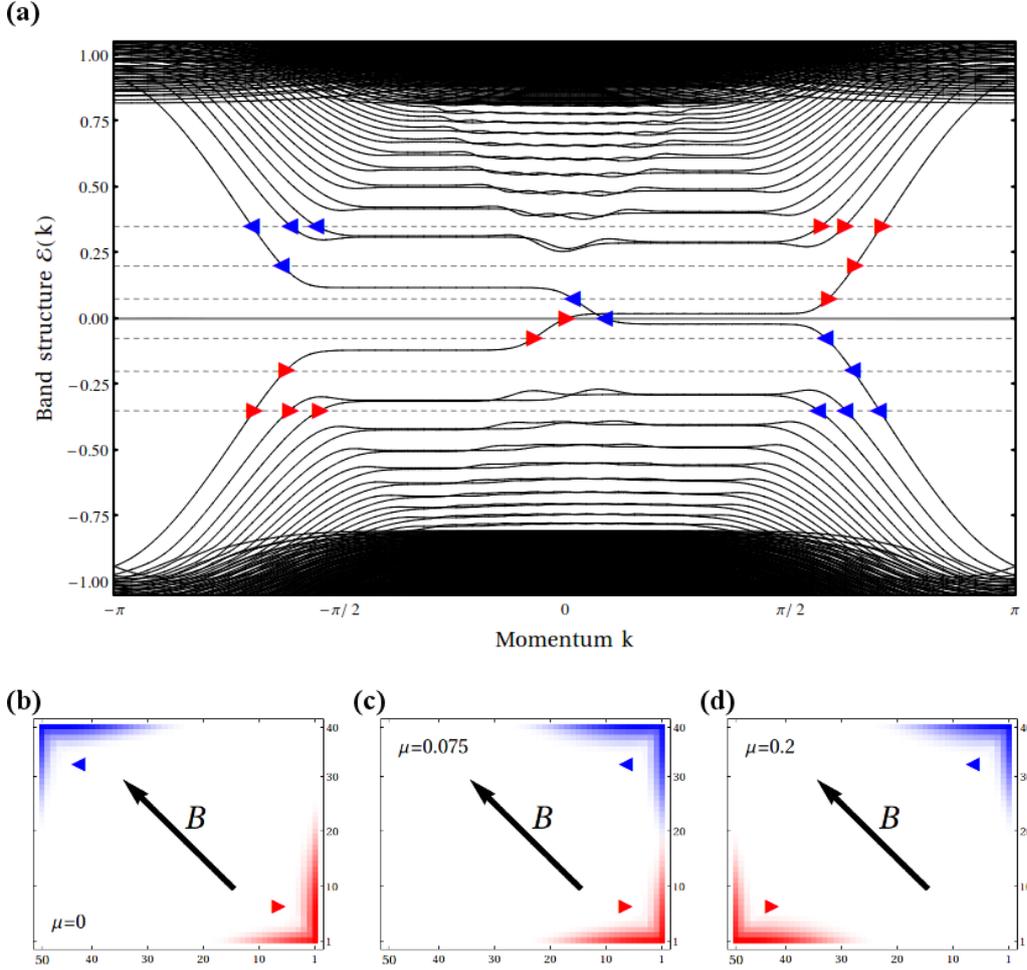


Figure 4.12. (a) Plot of the electronic band structure of a beam-shaped strong topological insulator with dimensions $N_x \times N_y \times N_z = \infty \times 50 \times 40$. The parameters of the model have been chosen such that it realized an inversion-asymmetric strong topological insulator in the absence of magnetic fields: $t = 1$, $m/t = 2$, $\Delta/t = 0.05$, $N_y = 50$, $N_z = 40$, $g_{\pm} = \frac{1}{2}$, $\mathbf{B}_Z = (0, 0.14, 0.14)^T$, $N_{\Phi}^y = N_{\Phi}^z = 0.007$. (b–d) Logarithmic plots of the probability distribution $|\psi(\mathbf{r})|^2$ as function of the coordinate \mathbf{r} in a cross section of the beam-shaped sample for different chemical potentials μ/t . Particles moving in positive (negative) x direction are denoted by a filled right (left) triangle in red (blue) color. For positive (negative) μ the edge channels are dominated by the orbital field and show an orbital motion in a plane perpendicular to the magnetic field, where the direction of the current-carrying edge channels is determined by the negative (positive) charge carriers. Panels (b–d) show an unusual sequence of Hall transitions, where the current-carrying edge channels move independently between the edges upon increasing (decreasing) the chemical potential μ/t . The presence of both magnetic Zeeman interaction and the inversion asymmetry Δ/t leads to a non-trivial interplay of the gaps induced by those interactions, resulting in a complicated behavior of the zeroth Landau level [cf. Eq. (4.48)] close to the Fermi level at $\mu/t = 0$ which is, however, also captured by our theory. We have published a similar figure in Ref. 19.

when changing or rotating the magnetic field, they could provide an interesting setup to study the interactions of individual edge channels.

4.2.2 Numerical results for counter-propagating edge channels

Finally, to gain a better understanding of the interaction effects of counter-propagating edge channels, which occur at step edges of the sample, we have calculated both the band structure and the corresponding eigenstates of the edge channels for various heights of the step edge. Fig. 4.13 shows the electronic band structure of beam with rectangular cross section, whose primary dimensions are given by $N_x \times N_y \times N_z = \infty \times 50 \times 40$. The step edge has a fixed width of 25 unit cells in y direction and a variable height in z direction, where panels (a) and (b) of Fig. 4.13 correspond to step heights of 1 and 20 unit cells, respectively. While the additional features in the electronic band structure due to the step edge are not very pronounced for small steps, they become more and more important for increasing step heights. In particular, we have calculated the eigenstates for a large step edge as illustrated in within the tight-binding approach for different chemical potential. As expected, we observe the current-carrying edge channels close to the edges, where the Hall quantum number ν changes. In Fig. 4.13, those states are the left- and right-moving states indicated by filled left (right) triangles with blue (red) color at energy $\mathcal{E}/t = -0.095$ and close to $k = 0$.

In principle, the tight-binding approach in conjunction with exact diagonalization methods would also allow to estimate the strength of the Coulomb interaction $V_{\text{Coul}}(\mathbf{r}, \mathbf{r}') \propto e^2/|\mathbf{r} - \mathbf{r}'|$ between the counter-propagating edge channels as well. Here, the field lines of the electric field can only extend in the yz plane, because we assume translational invariance in the direction perpendicular to the cross section, which also gives rise to the conserved momentum k in the band structure calculations. However, for a proper estimate of the Coulomb interaction strength one should also consider the dielectric properties of the model. For example, bulk HgTe has a large dielectric constant, $\epsilon_{\text{HgTe}} \approx 21.0$, and thus the screening of long-ranged Coulomb interactions is very efficient¹⁰². Moreover, due to the step edge the electric field shows a strong anisotropy within the 2D yz plane, so that an estimate of the Coulomb interaction would become much more complicated.

Finally, from a theoretical point of view it would also be desirable to investigate the interactions of counter-propagating edge channels in terms of chiral Luttinger liquids. Here, the general framework of Luttinger liquids should provide a suitable language to discuss the signatures of different types of interactions between counter-propagating chiral liquids.

4.3 The magneto-electric effect and the topological θ -term

In two and three dimensions, the non-interacting topological insulators are described by \mathbb{Z}_2 invariants (see sections 2.5 and 3.6.6) which have been reformulated in a number of different ways^{3,8,11,38–46}. However, in the presence of interactions or disorder, those expressions for the \mathbb{Z}_2 invariants are, in general, not well-defined. One approach to generalize the \mathbb{Z}_2 invariants to the interacting case can be formulated in a **topological field theory**¹¹. The relevant topological field theory in $3 + 1$ dimensions is given by the **axion Lagrangian** or the so-called **θ -term** which is well-known in quantum chromodynamics^{103–105}. Based on this term in the topological field theory Qi, Hughes, and Zhang have predicted, among other interesting effects, the so-called **topological magneto-electric effect**, where an electric field induces a magnetic field in the same direction. For example, a particularly interesting effect is the appearance of an image magnetic monopole under the surface of a topological insulator when an electric charge

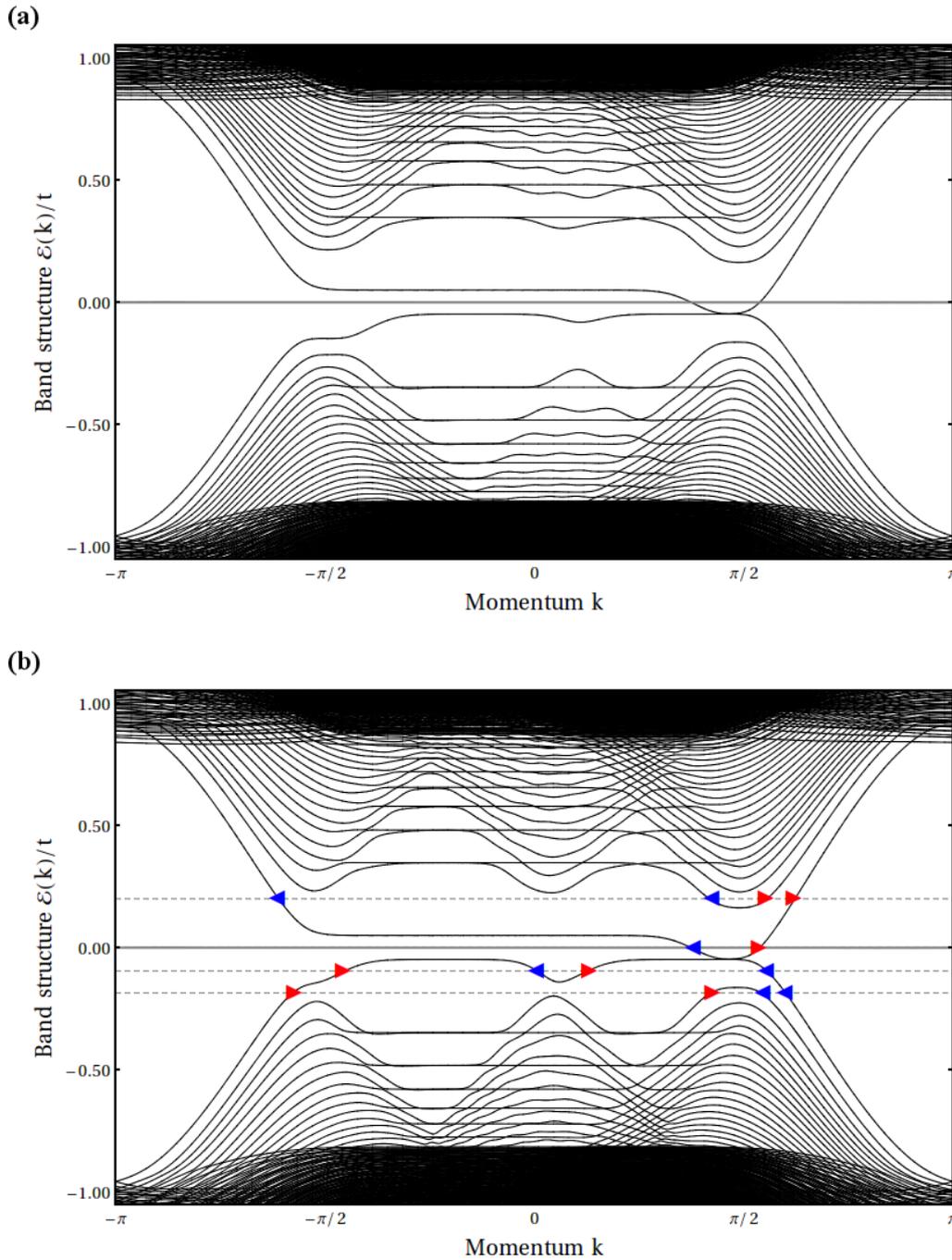


Figure 4.13. Plots of the electronic band structure of a beam-shaped strong topological insulator with dimensions $N_x \times N_y \times N_z = \infty \times 50 \times 40$. The parameters of the model have been chosen such that it realized an inversion-asymmetric strong topological insulator in the absence of magnetic fields: $t = 1$, $m/t = 2$, $\Delta/t = 0$, $N_y = 50$, $N_z = 40$, $g_{\pm} = \frac{1}{2}$. The magnetic field orientation has been chosen such that the topological insulator realizes the situation shown in Fig. 4.9. We have checked numerically that the edge channels are indeed located at the edges of the step edge sample. Panel (a) shows the band structure for a step edge with a height of two unit cells. Note that the additional features in the band structure close to $k = 0$ are not very pronounced, and fine-tuning of the magnetic field and/or the chemical potential is required to observe the counter-propagating edge channels. Moreover, due to the small height, the tunneling probability between those counter-propagating edge channels is quite large, so that tunneling leads to backscattering. In panel (b) we calculated the band structure for a step height of 20 unit cells. Here, the most interesting feature occurs at small negative chemical potential $\mu/t = -0.095$ close to $k = 0$, where the left- and right-moving states indicated by the filled left (right) triangle in blue (red) color correspond to the counter-propagating channels at the step edge.

is brought into proximity¹¹. The constant of proportionality entering the topological magneto-electric effect is universal and quantized in odd multiples of the fine-structure constant $\alpha = \frac{e^2}{\hbar c}$ (Ref. 11).

Let us first discuss the origin of the θ -term. As shown by Qi, Hughes, and Zhang, the effective action of a Dirac Hamiltonian in $3 + 1$ dimensions coupled to an external $U(1)$ gauge field $A_\mu(\mathbf{x}, t)$ can be derived by dimensional reduction². Starting from a Chern-Simons field theory in $4 + 1$ dimensions, one can choose a certain Landau gauge in which the Hamiltonian becomes translationally invariant in one of the spatial directions. Thus, assuming periodic boundary conditions in that direction, one can perform a Fourier decomposition of the spinors and rewrite the Hamiltonian as one-dimensional Bloch Hamiltonian. The effective action for the resulting system in $3 + 1$ dimensions can then be written as

$$\mathcal{S}_{\text{axion}}[A_\mu, \theta] = \frac{e^2}{\hbar} \frac{1}{4\pi} \int d^3x dt \epsilon_{\mu\nu\rho\sigma} \theta(\mathbf{x}, t) \partial_\mu A_\nu(\mathbf{x}, t) \partial_\rho A_\sigma(\mathbf{x}, t), \quad (4.54)$$

where $\mu, \nu, \rho, \sigma \in \{0, 1, 2, 3\} = \{t, x, y, z\}$ denote temporal and spatial coordinates, and $\epsilon_{\mu\nu\rho\sigma}$ is the totally anti-symmetric tensor with $\epsilon_{0123} = 1$. $\theta(\mathbf{x}, t)$ is a smooth field in the $3 + 1$ dimensional space-time, so that it can be considered locally as a parameter. This effective theory is known in the field theory literature as **axion electrodynamics**¹⁰³⁻¹⁰⁵, where the $\theta(\mathbf{x}, t)$ field plays the role of the axion field^{106,107}.

Usually, the electric polarization \mathbf{P} couples linearly to the external electric field \mathbf{E} , and the magnetization \mathbf{M} couples linearly to the magnetic field \mathbf{B} . However, the field $\theta(\mathbf{x}, t)$ couples to both external electromagnetic fields \mathbf{E} and \mathbf{B} since

$$\epsilon_{\mu\nu\rho\sigma} \partial_\mu A_\nu \partial_\rho A_\sigma = 2 \mathbf{E} \cdot \mathbf{B}. \quad (4.55)$$

For that reason, the θ -term is often referred to as **magneto-electric polarization**. Note that $\mathbf{E} \cdot \mathbf{B}$ is odd under time-reversal, so that in general the effective action $\mathcal{S}_{\text{axion}}$ breaks time-reversal symmetry. Note also that when $\theta(\mathbf{x}, t)$ becomes a constant parameter independent of space and time, *i.e.*, $\theta(\mathbf{x}, t) = \theta = \text{const}$, this effective action is also referred to as the topological term for the **θ -vacuum**^{108,109}. While this term is usually not accessible experimentally, Qi, Hughes, and Zhang have made several theoretical proposals, where the θ -field might be related to physical observables¹¹.

4.3.1 The value of θ in topological insulators

Let us start with a brief discussion of some properties related to the θ -term. First, if we consider the bulk of the system or a system with periodic boundary conditions in all spatial and temporal directions, $\theta(\mathbf{x}, t)$ takes a constant value θ . As a consequence, the effective action $\mathcal{S}_{\text{axion}}$ can be written as a total derivative:

$$\mathcal{S}_{\text{axion}}[A_\mu, \theta] = \frac{e^2}{\hbar} \frac{\theta}{4\pi} \int d^3x dt \epsilon_{\mu\nu\rho\sigma} \partial_\mu A_\nu \partial_\rho A_\sigma \quad (4.56a)$$

$$= \frac{e^2}{\hbar} \frac{\theta}{2\pi} \int d^3x dt \epsilon_{\mu\nu\rho\sigma} \partial_\mu (A_\nu \partial_\rho A_\sigma). \quad (4.56b)$$

At first glance, the above integral vanishes for a general electromagnetic field since it can be written as a surface integral in a closed $(3 + 1)$ -dimensional space-time, but this is misleading

²Note that this theory can also be obtained directly in a $3 + 1$ dimensional theory by integrating out the fermion fields in the presence of the external gauge fields $A_\mu(\mathbf{x}, t)$ and $\theta(\mathbf{x}, t)$.

as was pointed out, *e.g.*, by Vazifeh and Franz¹¹⁰. Since the gauge field $A_\mu(\mathbf{x}, t)$ of a periodic electromagnetic field is, in general, not periodic in space and time, the effective action $\mathcal{S}_{\text{axion}}$ does not vanish. Instead, the integral is rather quantized as $2\pi m$ with integer m in units of e^2/h (Ref. 110), so that the action $\exp(i\mathcal{S}_{\text{axion}}) = \exp(i\theta m)$ is invariant under the shift of θ by any integer multiple of 2π :

$$\exp(i\theta m) = \exp[i(\theta + 2\pi n)m] \quad \forall n \in \mathbb{Z}. \quad (4.57)$$

All physical observables are thus invariant under a global transformation $\theta \rightarrow \theta + 2\pi n$ with integer n . Moreover, in the presence of time-reversal symmetry ($\hat{\Theta}$) and/or spatial inversion symmetry ($\hat{\mathcal{P}}$), the electromagnetic fields \mathbf{E} and \mathbf{B} transform as

$$\hat{\Theta}\mathbf{E} = \mathbf{E}, \quad \hat{\Theta}\mathbf{B} = -\mathbf{B}, \quad (4.58a)$$

$$\hat{\mathcal{P}}\mathbf{E} = -\mathbf{E}, \quad \hat{\mathcal{P}}\mathbf{B} = \mathbf{B}. \quad (4.58b)$$

Both symmetries independently imply that the product $\mathbf{E} \cdot \mathbf{B}$ is odd, *i.e.*, changes sign under $\hat{\Theta}$ or $\hat{\mathcal{P}}$. Thus, naively also θ is odd, and the only allowed value for θ is zero. This, however, neglects the fact that the action is also 2π -periodic, and shifting θ by multiples of 2π shows that also $\theta = \pi$ is allowed by the above symmetries. As a consequence, in a time-reversal invariant and/or inversion-symmetric system, the value of θ must be equal to either $\theta = 0$ or $\theta = \pi$. It was believed for a long time, however, that in most insulating materials $\theta = 0$ is the only allowed value for the θ -term due to the lack of experimental evidence for systems realizing the non-trivial value of $\theta = \pi$. Nevertheless, the recent discovery of the 3D topological insulators and the quantization of the θ value allow for a distinction between strong topological insulators and weak topological insulators or trivial band insulators, as was pointed out by Qi, Hughes, and Zhang¹¹:

$$\frac{\theta}{2\pi} = \begin{cases} n & \text{for weak topological insulators and trivial band insulators} \\ n + \frac{1}{2} & \text{for strong topological insulators} \end{cases} \quad (4.59)$$

Finally, note that if both time-reversal and inversion symmetry are broken, the topological quantization of physical response functions is not guaranteed any longer which implies that the value of θ is not quantized any more and can take any fractional value. In that case, one has to use other methods to define and compute the topological invariant ν_0 , for example Green's functions.

4.3.2 Relation between the θ -term and the Hall response

As previously mentioned, due to the 2π periodicity of the effective action $\mathcal{S}_{\text{axion}}$, *i.e.*, $\theta \rightarrow \theta + 2\pi n$, the integer part of the $\frac{\theta}{2\pi}$ is not a physical quantity. However, as we discuss below, if we consider a system with open boundary conditions, the integer part of θ entering the effective action $\mathcal{S}_{\text{axion}}$ becomes physical, and its value depends on quantitative details of the boundary.

Being quadratic in the external gauge field A_μ , the effective action $\mathcal{S}_{\text{axion}}$ describes the linear response of the system to external electromagnetic fields \mathbf{E} and \mathbf{B} . By taking the functional derivative of $\mathcal{S}_{\text{axion}}$ with respect to A_μ we obtain the electromagnetic response of the system in terms of the generalized current j_μ which depends on the spatial and temporal gradients of the θ -field:

$$j_\mu(\mathbf{x}, t) \equiv \frac{\delta \mathcal{S}_{\text{axion}}[A_\mu, \theta]}{\delta A_\mu(\mathbf{x}, t)} = -\frac{e^2}{h} \frac{1}{2\pi} \epsilon_{\mu\nu\rho\sigma} \partial_\nu \theta(\mathbf{x}, t) \partial_\rho A_\sigma(\mathbf{x}, t). \quad (4.60)$$

In other words, a temporal or spatial gradient of $\theta(\mathbf{x}, t)$ induces a Hall current in the system. Conversely, there will be no Hall response in the bulk, where $\theta(\mathbf{x}, t)$ takes a constant value θ and where all gradients of $\theta(\mathbf{x}, t)$ vanish.

However, if we consider a system with an interface between two homogeneous materials, where the θ -field depends only on the z coordinate and is constant otherwise, *i.e.*, $\theta(\mathbf{x}, t) = \theta(z)$, then the current density is given by

$$j_\mu(z) = -\frac{e^2}{h} \frac{\epsilon_{\mu\nu\sigma}}{2\pi} \partial_z \theta(z) \partial_\nu A_\sigma(\mathbf{x}, t) \quad (4.61)$$

with $\mu, \nu, \sigma \in \{t, x, y\}$. In the static limit, this current describes the quantum Hall effect in the perpendicular xy plane with the corresponding Hall conductivity $\sigma_{xy}(z)$ given by [see also Fig. 4.14 (a)]:

$$\sigma_{xy}(z) \equiv \frac{e^2}{h} \frac{\partial_z \theta(z)}{2\pi}. \quad (4.62)$$

By integrating $\sigma_{xy}(z)$ in a finite range $z_1 \leq z \leq z_2$ we obtain the interface current as

$$J_y \equiv \int_{z_1}^{z_2} dz j_y(z) = \int_{z_1}^{z_2} dz \sigma_{xy}(z) E_x = \frac{e^2 E_x}{h} \frac{1}{2\pi} \int_{z_1}^{z_2} dz \partial_z \theta(z) = \frac{e^2 E_x}{h} \frac{1}{2\pi} \int_{\theta(z_1)}^{\theta(z_2)} d\theta. \quad (4.63)$$

Thus, the Hall conductance of the region between z_1 and z_2 depends only on the change in $\theta(z)$ in this region and is not sensitive to any details of the function $\theta(z)$:

$$\sigma_{xy}^{\text{surf}} = \frac{e^2}{h} \int_{z_1}^{z_2} dz \frac{\partial_z \theta(z)}{2\pi} = \frac{e^2}{h} \int_{\theta(z_1)}^{\theta(z_2)} \frac{d\theta}{2\pi} = \frac{e^2}{h} \frac{\theta(z_2) - \theta(z_1)}{2\pi}. \quad (4.64)$$

As a result, an interface between the two homogeneous materials with different θ values has the Hall conductance

$$\sigma_{xy}^{\text{surf}} = \frac{e^2}{h} \frac{\Delta\theta}{2\pi}, \quad (4.65)$$

where $\Delta\theta = \theta(z_2) - \theta(z_1)$ denotes the difference of the θ values of both materials. More importantly, since $\theta = \pi$ in a strong topological insulator, we observe that the Hall conductance at the interface between a trivial band insulator and the strong topological insulator is equal to half of the conductance quantum:

$$\sigma_{xy}^{\text{surf}} = \pm \frac{e^2}{2h}. \quad (4.66)$$

Note that this result is valid only if the surface band structure is gapped, for example, by the proximity to a ferromagnetic layer.

This result can also be understood as follows: The effective Hamiltonian for a single Dirac fermion at the interface can always be written in a proper basis as two-band Bloch Hamiltonian

$$\mathcal{H}_{\text{surf}}(k_x, k_y) = v_F(k_x \sigma_x + k_y \sigma_y) + m \sigma_z, \quad (4.67)$$

where m is an effective mass which is, for example, generated due to the proximity to a ferromagnetic layer. This Dirac Hamiltonian can also be obtained, for example, by a Taylor expansion of the Haldane model introduced in chapter 2 close to the quantum phase transition at $m/t = 2$. As shown for the Haldane model, one can compute the Hall conductance from the Chern number which is well-defined if the fermion mass m is non-zero so that a band gap is

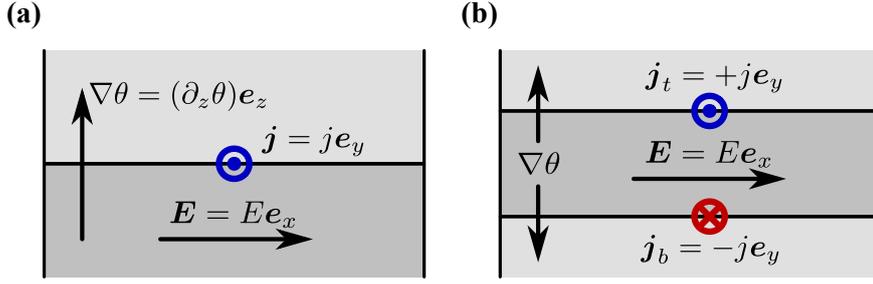


Figure 4.14. (a) Illustration of the Hall effect at a single interface between two homogeneous materials induced by the spatial gradient $\partial_z\theta(z)$. (b) Illustration of the Hall effect at the two parallel interfaces in a heterostructure. The applied electric field $\mathbf{E} = E\mathbf{e}_x$ induces Hall currents \mathbf{j}_t and \mathbf{j}_b at the top and bottom interfaces of the topological insulator (TI) and the trivial band insulator (BI), where the θ value changes. These Hall currents \mathbf{j}_t and \mathbf{j}_b point into and out of the direction of the paper, respectively.

opened. For a constant mass m it turns out that the Hall conductance for a single Dirac fermion is then equal to a half of the conductance quantum:

$$\sigma_{xy}^{2D} = \frac{e^2}{2h} \text{sgn}(m). \quad (4.68)$$

Quite generally, the topology of the bulk band structure requires that the Hall conductance carried by a single surface Dirac fermion at the interface between two homogeneous materials with different θ values is quantized in half-odd integers, provided a surface band gap exists. Note that the half-integer quantization of the Hall conductance of a single Dirac fermion is intimately related to its π Berry phase^{11,28,95,96}. In that sense, the above result for the surface conductance $\sigma_{xy}^{\text{surf}}$ is intimately related to the Hall conductance in the Haldane model.

Aside from the unusual half-integer quantization of the Hall conductance, Eq. (4.66) seems to violate Laughlin's quantization of the integer quantum Hall effect²². The resolution to this apparent contradiction is that in a real sample there are always two surfaces at opposite sides of the sample with opposite normal vectors. As a consequence, the total Hall response $\sigma_{xy}^{\text{slab}} = \sigma_{xy,t}^{\text{surf}} - \sigma_{xy,b}^{\text{surf}}$ in a slab with top and bottom surface is either equal to 0 or ± 1 , and thus an integer number, thereby restoring the integer quantization of the quantum Hall effect.

To clarify the consequences of the last point, let us consider a topological insulator in a slab geometry which is sandwiched on both sides by trivial band insulators [see Fig. 4.14 (b)]. Furthermore, we assume that time-reversal symmetry is broken by an applied magnetic field which creates a band gap in the surface electronic band structure. As shown above, the topological insulator is characterized by a non-trivial value of θ which, in the presence of time-reversal or inversion symmetry, takes the bulk value $\theta = (2n+1)\pi$ with integer n . The trivial band insulators, on the other hand, shall be described by the values $\theta_{t,b} = 2\pi n_{t,b}$ for the top and bottom layers, respectively. Due to the spatial gradient in $\theta(z)$ across the interfaces, the applied electric field $\mathbf{E} = E\mathbf{e}_x$ induces Hall currents in the y direction given by [cf. Eqs. (4.60, 4.61)]:

$$\mathbf{j}_t = -\frac{e^2}{2\pi h} (\nabla\theta) \times \mathbf{E} = +\frac{e^2}{h} \frac{\theta_t - \theta}{2\pi} E_x \mathbf{e}_y, \quad (4.69a)$$

$$\mathbf{j}_b = -\frac{e^2}{2\pi h} (\nabla\theta) \times \mathbf{E} = -\frac{e^2}{h} \frac{\theta_b - \theta}{2\pi} E_x \mathbf{e}_y. \quad (4.69b)$$

The corresponding Hall conductances for the individual top and bottom interface are then given

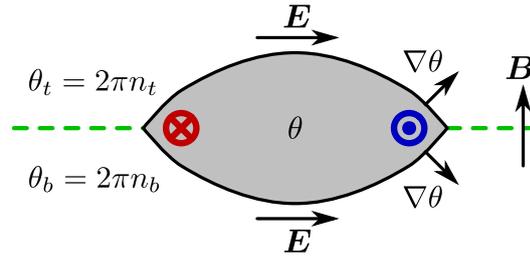


Figure 4.15. Schematic picture of the quantum Hall effect at the edges of a topological insulator, with the dashed green lines indicating the branch cuts of θ separating the top and bottom vacuum from each other. While the spatial gradient $\partial_z\theta(z)$ induces a finite surface Hall response $\sigma_{xy}^{\text{surf}}$ at the top and bottom interfaces, the orbital magnetic field B “pushes” the induced Hall currents to the edges of the sample. At the edges, where the top and bottom surfaces intersect each other, the total Hall response $\sigma_{xy}^{\text{Hall}} = \sigma_{xy,t}^{\text{surf}} - \sigma_{xy,b}^{\text{surf}} = \frac{e^2}{h}(n_t - n_b)$ is quantized and independent of the precise value of θ within the bulk of the topological insulator.

by

$$\sigma_{xy,t}^{\text{surf}} = \frac{e^2}{h} \frac{\theta_t - \theta}{2\pi} = \frac{e^2}{h} \left(n_t - \frac{\theta}{2\pi} \right), \quad (4.70a)$$

$$\sigma_{xy,b}^{\text{surf}} = \frac{e^2}{h} \frac{\theta_b - \theta}{2\pi} = \frac{e^2}{h} \left(n_b - \frac{\theta}{2\pi} \right). \quad (4.70b)$$

Apparently, the individual surfaces of a strong topological insulator with $\theta = \pi$ show a half-integer quantized Hall conductance. However, if we consider that the top and bottom surfaces intersect at some point in real space (or infinity for that matter), then the total Hall response of the topological insulator at this edge is given by the difference of the surface Hall conductances since the surface normal vectors point in opposite directions:

$$\sigma_{xy}^{\text{Hall}} = \sigma_{xy,t}^{\text{surf}} - \sigma_{xy,b}^{\text{surf}} = \frac{e^2}{h}(n_t - n_b). \quad (4.71)$$

As a result, we obtain the well-known integer-quantized Hall effect: $\sigma_{xy}^{\text{Hall}} = n(e^2/h)$ with $n \in \mathbb{Z}$. Most importantly, the total Hall response at the intersection is independent of the particular value of θ in the bulk of the topological insulator. Thus, the precise value of θ cannot be obtained from a standard quantum Hall measurement probing $\sigma_{xy}^{\text{Hall}}$, because a transport measurement is but a global measurement probing the whole topological insulator sample. Nevertheless, by using appropriate surface-sensitive methods one could, in principle, determine the value of $\theta \bmod 2\pi$ from $\sigma_{xy}^{\text{surf}}$ [cf. Eq. (4.66)], because they provide access to local properties of the sample. For example, the polar Kerr effect, where linearly polarized light shows a rotation of the polarization axis upon reflection on some surface, is a well-known experimental probe of the Hall conductance σ_{xy} , and it has been recently applied to thin films of topological insulators in the vicinity of a ferromagnet^{111,112}.

Moreover, we immediately see from Eq. (4.71) that the appearance of current-carrying edge states implies that the top and bottom vacuum surrounding the topological insulator are different in the sense that $n_t \neq n_b$. Put differently, the value of θ is not well-defined throughout the whole space, but contains singular branch cuts, where the integer part of $\frac{\theta}{2\pi}$ changes (see Fig. 4.15). As a consequence, the singularities of the θ -term and the chiral edge states are also related to one another.

To better understand those chiral edge states, let us first compute the currents through a 2D manifold with an edge-like boundary (see Fig. 4.16). Similar to a topological insulator in the slab geometry, the planar surfaces of the crystal can be characterized by integers n_1 and n_2 describing the integer part of the surface Hall response $\sigma_{xy}^{\text{surf}}$ [cf. Eq. (4.70)]. In the static limit, i.e., $\partial_t \theta(\mathbf{r}, t) = 0$, the current density $\mathbf{j} = j e_y$ as obtained from Eq. (4.60) can be rewritten as

$$\mathbf{j}(\mathbf{r}) = -\frac{e^2}{2\pi h} (\nabla \theta(\mathbf{r})) \times \mathbf{E}(\mathbf{r}). \quad (4.72)$$

By integrating this current density over a sufficiently large region across the edge bounded by the contour \mathcal{C} (see Fig. 4.16), we obtain the following current-carrying edge channel:

$$I = \int d\sigma \hat{\mathbf{n}} \cdot \mathbf{j}(\mathbf{r}) = -\frac{e^2}{2\pi h} \int d\sigma \hat{\mathbf{n}} \cdot (\nabla \theta(\mathbf{r})) \times \mathbf{E}(\mathbf{r}) = -\frac{e^2}{2\pi h} \int d\sigma \hat{\mathbf{n}} \cdot \nabla \times (\theta(\mathbf{r}) \mathbf{E}(\mathbf{r})). \quad (4.73)$$

Since the electric field is a conservative force and can thus be described in terms of a scalar potential $\Phi(\mathbf{r})$ with $\mathbf{E}(\mathbf{r}) = -\nabla \Phi(\mathbf{r})$, where $\nabla \times \mathbf{E}(\mathbf{r}) = 0$, we can write the above equation for the current I as

$$I = -\frac{e^2}{2\pi h} \oint_{\mathcal{C}} d\mathbf{r} \cdot (\theta(\mathbf{r}) \mathbf{E}(\mathbf{r})) = \frac{e^2}{2\pi h} \oint_{\mathcal{C}} d\mathbf{r} \cdot \theta(\mathbf{r}) (\nabla \Phi(\mathbf{r})) = -\frac{e^2}{2\pi h} \oint_{\mathcal{C}} d\mathbf{r} \cdot (\nabla \theta(\mathbf{r})) \Phi(\mathbf{r}), \quad (4.74)$$

where we have rewritten the surface integral of the curl of the vector field $\theta(\mathbf{r}) \mathbf{E}(\mathbf{r})$ into a line integral along the boundary \mathcal{C} of the surface by applying Stokes' theorem. If we assume further that $\theta(\mathbf{r})$ is singular at most at one point on the manifold, we may rewrite the expression for the current as

$$I = -\frac{e^2}{2\pi h} \Delta \Phi \oint_{\mathcal{C}} d\mathbf{r} \cdot \nabla \theta(\mathbf{r}) = \frac{e^2}{h} \Delta \Phi (n_1 - n_2), \quad (4.75)$$

where $\Delta \Phi \equiv \Phi_e - \Phi_s$ denotes the potential difference between the edge equipotential line with $\Phi(\mathbf{r}) = \Phi_e$ and the potential Φ_s at the singularity (dashed green line in Fig. 4.16). Therefore, the number of one-dimensional chiral edge channels is given by an integer winding number \mathcal{W} of the θ -term, and the Hall conductance $\sigma_{xy}^{\text{Hall}} = I/\Delta \Phi$ can be rewritten as

$$\sigma_{xy}^{\text{Hall}} = -\frac{e^2}{h} \mathcal{W} \quad \text{with} \quad \mathcal{W} \equiv \frac{1}{2\pi} \oint_{\mathcal{C}} d\mathbf{r} \cdot \nabla \theta(\mathbf{r}) = n_1 - n_2. \quad (4.76)$$

This underlines the quantization of the Hall conductance in units of e^2/h . Moreover, the location of the edge channels tracks the singularities of the θ -term and vice versa. Since the winding numbers \mathcal{W} are integer-quantized and thus independent of the precise value of θ characterizing the topological insulator, the total Hall response $\sigma_{xy}^{\text{Hall}}$ is also independent of details of θ . This shows that the quantization of the Hall effect, rooted in charge quantization, is a much more robust concept than the quantization of the θ value in topological insulators which is based on the presence of time-reversal symmetry or spatial inversion symmetry.

4.3.3 Numerical results for the θ -term

Finally, for an infinite slab in the xy plane with dimensions $N_x \times N_y \times N_z = \infty \times \infty \times 20$ we have also calculated the layer-resolved matrix of conductivities $\sigma_{xy}(z, z')$ and derived the change of the θ -term as function of the layer index z . Specifically, we consider the following

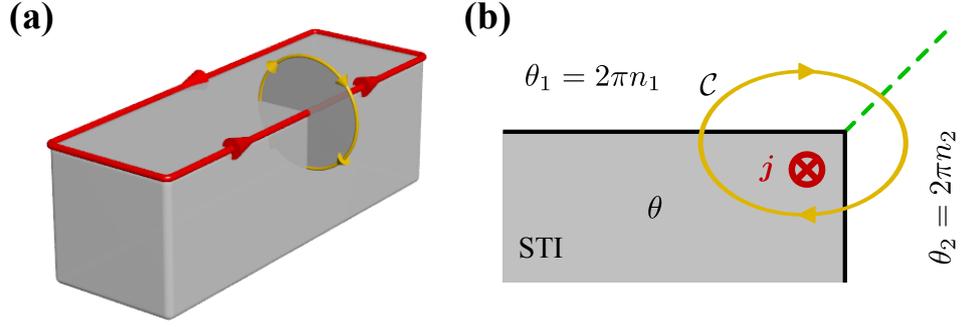


Figure 4.16. **a)** Illustration of a current-carrying edge channel running along the edges of a topological insulator. **b)** Schematic view in the direction of the current. The planar surfaces of the topological insulator are characterized by integers n_1 and n_2 , describing the integer change of the θ value across the surfaces, thereby characterizing the surface Hall response σ_{xy}^{2D} [cf. Eq. (4.70)]. The number of chiral edge channels can be obtained by computing the winding number of the θ -term along the contour \mathcal{C} in a clockwise orientation. The dashed green line indicates the branch cut in real space, where $\theta(\mathbf{r})$ is singular. We have published parts of this figure in Ref. 19.

one-dimensional tight-binding Hamiltonian which is parametrized by a two-dimensional in-plane momentum $\mathbf{k} = (k_x, k_y)^T$:

$$\begin{aligned}
\mathcal{H}(\mathbf{k}) = & \sum_z \Psi_{\mathbf{k}z}^\dagger (m - t \cos(k_x a) - t \cos(k_y a)) \Gamma_0 \Psi_{\mathbf{k}z} \\
& - t \sum_z \Psi_{\mathbf{k}z}^\dagger (\sin(k_x a) \Gamma_1 + \sin(k_y a) \Gamma_2) \Psi_{\mathbf{k}z} \\
& - t \sum_z \left[\Psi_{\mathbf{k}z}^\dagger \left(\frac{\Gamma_0 - i\Gamma_3}{2} \right) \Psi_{\mathbf{k},z-a}^\dagger + \text{H.c.} \right] \\
& + \Delta \sum_z \Psi_{\mathbf{k}z}^\dagger \Gamma_4 \Psi_{\mathbf{k}z} + \mathcal{H}_Z.
\end{aligned} \tag{4.77}$$

Here, the first two lines describe the usual hopping of electrons within a single layer of the slab, whereas the third line describes the inter-layer hopping between two adjacent layers. We also include a Zeeman interaction \mathcal{H}_Z which creates a finite bulk band gap in the electronic band structure, and thus the energy denominator entering the Kubo formula (4.78) below becomes well-defined. The layer-resolved Hall response can then be calculated from the matrix of conductivities $\sigma_{xy}(z, z')$ using the Kubo formula³³ as follows:

$$\begin{aligned}
\sigma_{xy}(z, z') = & i \frac{e^2}{h} \sum_{\substack{\alpha, \beta \\ \mathcal{E}_\alpha < \mu < \mathcal{E}_\beta}} \int d^2k \left[\frac{\langle \alpha | \partial_{k_x} \mathcal{H}(\mathbf{k}, z) | \beta \rangle \langle \beta | \partial_{k_y} \mathcal{H}(\mathbf{k}, z') | \alpha \rangle}{(\mathcal{E}_\alpha - \mathcal{E}_\beta)^2} \right. \\
& \left. - \frac{\langle \alpha | \partial_{k_y} \mathcal{H}(\mathbf{k}, z') | \beta \rangle \langle \beta | \partial_{k_x} \mathcal{H}(\mathbf{k}, z) | \alpha \rangle}{(\mathcal{E}_\alpha - \mathcal{E}_\beta)^2} \right], \tag{4.78}
\end{aligned}$$

where $\mathcal{H}(\mathbf{k}, z)$ denotes the projection of the above tight-binding Hamiltonian $\mathcal{H}(\mathbf{k})$ onto layer z . Consequently, the total response $\sigma_{xy}(z)$ in layer z to a uniform electric field is given by:

$$\sigma_{xy}(z) = \int_{-\infty}^{\infty} dz' \sigma_{xy}(z, z'), \tag{4.79}$$

which allows us to calculate the change in the θ -term from one layer to next one as follows:

$$\theta(z) = \int_{-\infty}^z dz' \sigma_{xy}(z'). \quad (4.80)$$

Fig. 4.17 (a, b) shows the layer-resolved Hall response $\sigma_{xy}(z)$ (top panel) and the integrated value for the θ -term (bottom panel) as function of the layer index z for different phases realized by the minimal model. First, note that in the trivial band insulator phase, *i.e.*, for $|m/t| > 3$, the system shows only a very small transverse response at all which is in agreement with our naive understanding of the trivial band insulating character of this phase. Consequently, there is no change in the value of $\theta(z)$ across the slab. The situation is, however, completely different in the topologically non-trivial phases for $|m/t| < 3$. Both weak and strong topological insulators are characterized by non-vanishing Hall responses at the top and bottom surfaces of the slab, where the gradient $\partial_z \theta(z)$ is largest. Nevertheless, we observe an important difference between the weak and strong topological insulator: While the weak topological insulator for $|m/t| < 1$ shows a Hall response $\sigma_{xy} \approx e^2/h$, the strong topological insulator for $1 < |m/t| < 3$ is characterized by half a conductance quantum $\sigma_{xy} \approx e^2/(2h)$ as expected from theory. In other words, the strong topological insulator hosts a single surface state at each surface, whereas the weak topological insulators hosts two surface states at each surface. Upon integrating the layer-resolved Hall response we immediately see that the value of the θ -term within the bulk of the slab is given by an even (odd) multiple of π in the weak (strong) topological insulator, thus underlining the formulation of the \mathbb{Z}_2 classification of time-reversal invariant and/or inversion-symmetric 3D topological insulators discussed in section 4.3.1.

However, as mentioned before, this \mathbb{Z}_2 classification in terms of θ is only well-defined in the presence of either time-reversal symmetry or inversion symmetry, because the value of θ is not quantized any longer if both symmetries are absent. To illustrate this further, we have calculated the θ -term of the minimal model (4.78) also in the presence of symmetry-breaking terms, as shown in Fig. 4.17 (c, d). To be specific, in the minimal model, such a symmetry breaking can be easily taken into account by considering a finite value of $\Delta \Gamma_4$. Starting from a time-reversal invariant and inversion-symmetric model for $\Delta/t = 0$ with quantized Hall response $\sigma_{xy} = e^2/(2h)$ at the surfaces and a quantized bulk value $\theta = \pi$, we observe that the bulk value of the θ -term is no longer quantized, but takes an arbitrary, non-quantized value when Δ/t is increased. Upon further increasing Δ/t beyond a critical value given by $\Delta_c/t = g_+ B_Z^z$, a surface quantum phase transition takes place, where the total Hall response of the slab $\sigma_{xy,t}^{\text{surf}} + \sigma_{xy,b}^{\text{surf}}$ changes from e^2/h to zero. This is accompanied by a sign change of the surface Hall conductance of one of the two surfaces and can be understood in terms of the change of the θ value in the bulk as follows. For $\Delta/t = 0$, one can characterize the top and bottom vacua surrounding the sample by different θ -values $\theta_t = 2\pi n_t$ and $\theta_b = 2\pi n_b$, where $n_t \neq n_b$ in the topologically non-trivial phase. When Δ/t is increased, both time-reversal and inversion symmetry are broken, so that the half-integer quantized surface Hall response of a 3D topological insulator is no longer protected, which can also be seen in the classification table of topological insulators^{65–67}. Indeed, we find a crossover of the θ -term in the *bulk* of the system, and at the critical value Δ_c the value of θ crosses an integer multiple of 2π which leads to the sign change in the surface Hall conductance and marks the surface quantum phase transition to the “trivial” phase.

To summarize those results, we see that the total Hall response and the total change in the θ -value across the whole sample are always quantized due to the topological value of $\theta_{t,b} = 2\pi n_{t,b}$ describing the vacua on either side of the slab, irrespective of the presence or absence of time-reversal symmetry or inversion symmetry. As a consequence, a quantum Hall experiment

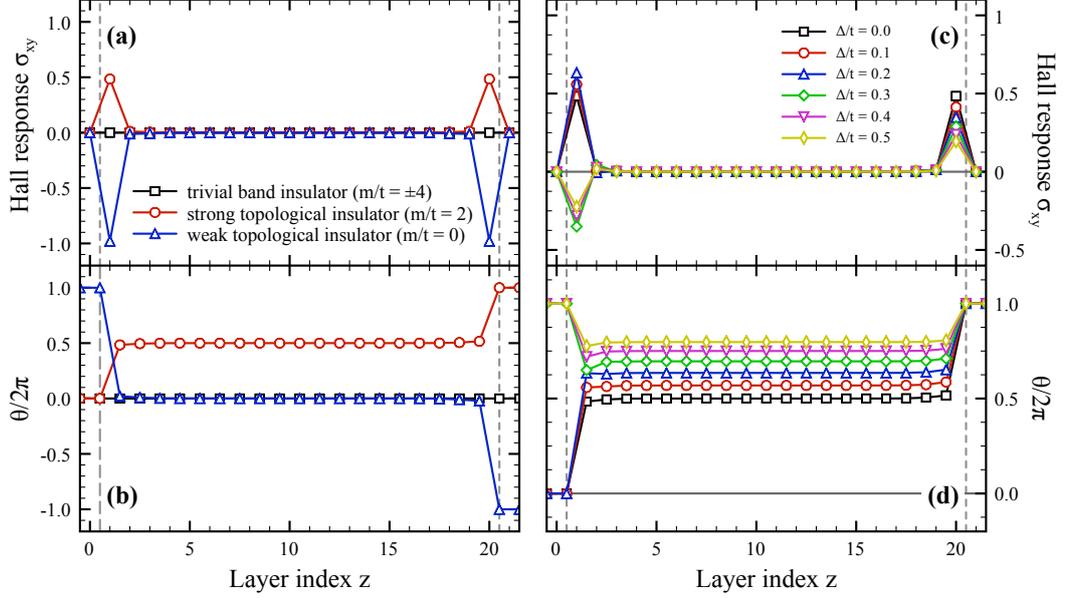


Figure 4.17. Plots of the layer-resolved in-plane Hall response $\sigma_{xy}(z)$ [panels (a) and (c)] and plots of the θ -angle [panels (b) and (d)] as function of the layer index z , calculated for different phases from a layer-resolved Kubo formula in a slab geometry for $N_z = 20$ layers [cf. Eqs. (4.78–4.80)]. The parameters of the model ($t = 1$, $\Delta/t = 0$, $g_+ B_z^z = 0.25$) are chosen such that it realizes a trivial band insulator phase for $m/t = \pm 4$ (black line, open squares), a strong topological insulator phase for $m/t = 2$ (red line, open circles), and a weak topological insulator phase for $m/t = 0$ (blue line, open triangles). (a) In the trivial insulator phase for $|m/t| > 3$ and $\Delta/t = 0$, there is almost no Hall response to an applied external electric field, while in the topological insulator phases for $|m/t| < 3$ both top and bottom surface host metallic surface states. (b) $\theta(z)$ is constant within the bulk of the sample, but changes when entering or leaving the topological insulator sample. Moreover, since $\theta(z)$ tracks the metallic surface states, we observe that the strong topological insulator hosts a single topological state on each surface, while the weak topological insulator has two topological surface states. The integer part of $\theta/(2\pi)$ has been chosen such that $0 \leq \theta \leq 2\pi$ in the bulk of the sample. (c) At $\Delta/t = g_+ B_z^z = 0.25$ the bulk of the system shows a crossover to a trivial band insulator phase which can be observed by the sign change of the surface Hall response at this point. (d) Upon increasing Δ/t we observe that the bulk value of θ is not an integer multiple of π , but takes any value, as expected from theory. Nevertheless, the total change across the sample remains quantized, $\theta(N_z) - \theta(1) = 2\pi(n_t - n_b)$. Note that in this case the strong topological insulator phase for $\Delta/t < 0.25$ is characterized by vacua with different topological quantum numbers $n_{t,b} = 0, 1$ on both sides of the slab, while the trivial band insulator has the same vacuum surrounding it. We have published a similar Figure in Ref. 19.

probing the whole slab can only measure the integer quantization of the Hall conductance $\sigma_{xy}^{\text{Hall}}$ which is in agreement with the fact that charge is quantized in multiple of the unit charge e . Thus, the value of the bulk θ -term is inaccessible to a transport measurement and can only be revealed in different types of surface-sensitive experiments. As mentioned before, the Kerr effect, where linearly polarized light shows a rotation of the polarization axis upon reflection on some surface, is a well-known experimental probe of the Hall conductance which could, in principle, measure the local value of $\sigma_{xy}^{\text{surf}}$.

4.4 Summary and discussion

In this chapter, we discussed important aspects of our minimal model for topological insulators and focused in particular on the quantum Hall effect which takes place on the 2D surfaces of a 3D topological insulator.

When deriving the effective Dirac Hamiltonian for a single surface we observed that the Zeeman interaction coupling the electron's spin to the magnetic field leads to a non-trivial and surface-dependent mass term for the surface Dirac fermions. We showed analytically that the competition between the gaps induced by the symmetry-breaking term $\Delta \Gamma_4$ and the Zeeman interaction leads to different types of edge channels which are either dominated by the orbital magnetic field or by the Zeeman field. In finite-size geometries we showed that those edge channels form networks since all surfaces are connected to each other. Moreover, upon changing the orientation of the magnetic field the Dirac mass term on the surfaces changed, and this results in non-trivial edge channel networks with two or more edge channels along a single edge. In particular, we showed that the corresponding conductance network also allows to measure fractional values of the conductance quantum when contacting the individual edges by voltages probes. Since this pattern changes as function of the magnetic field and/or the surface chemical potential, this allows for interesting experiments on 3D topological insulators. To illustrate the Landau level sequence in a magnetic field and the positions of the edge channels, we calculated the electronic band structure of a infinite beam with rectangular cross section. The energy of the resulting Landau levels agreed with the analytical sequence within a few percent, and also the edge channels were located within a few lattice sites at the edges of the sample. While the complicated interplay between the different gaps is more difficult to grasp analytically, the numerical tight-binding calculations allowed to study the behavior of the edge channels when changing, for example, the chemical potential. In particular, we observed edge channels "jumping" from one edge to an adjacent one, which could be used as a hint for the topological nature of a sample. Furthermore, the dependence on the relative orientation between magnetic field and surfaces allows for non-trivial conductance networks which can also show fractional conductances when placing appropriate voltage contacts on the edges.

In the second part of this chapter, we discussed the role of the θ -term in topological insulators and showed how it relates to the Hall response measured in transport experiments. Notably, we showed that the Hall response $\sigma_{xy}^{\text{Hall}}$ is a global measurement on the topological insulator and quantized in integer multiples of the conductance quantum e^2/h due to charge quantization, while in a time-reversal invariant topological insulator the response of a single surface, $\sigma_{xy}^{\text{surf}}$, is equal to half the conductance quantum. Nevertheless, this half-integer quantization ceases to exist once time-reversal and inversion symmetries are broken in the bulk of the sample, for example by the applied magnetic field, which was confirmed by numerically calculating the corresponding Hall response of a topological insulator slab.

The results of this chapter have been published in Ref. 19.

CHAPTER 5

Surface Criticality in 3D Topological Insulators

Non-interacting topological insulators are characterized by a bulk band gap and gapless edge or surface states at their boundaries which are protected by time-reversal symmetry and the intrinsic topological winding of the electronic bulk band structure, as discussed before. Therefore, topological insulators are stable against non-magnetic disorder or perturbations which do not break the symmetries of the Hamiltonian, and the bulk-boundary correspondence suggests that, as long as the bulk band structure remains gapped with a non-trivial band topology, there exist edge states at the boundaries. If an energy gap can be induced in the surface states, even more interesting effects such as the **topological magneto-electric effect** or **Majorana fermions** are expected to occur on the surfaces of topological insulators. This gap can be induced, for example, by breaking time-reversal symmetry in the bulk by applying an external magnetic field⁸ (compare also previous chapter), by proximity to a magnetic material¹¹ or a superconductor¹¹³, etc. However, it is not immediately clear to what extent the bulk-boundary correspondence holds in the presence of electron-electron interactions. In general, interactions are expected to create correlated topological states¹¹⁴⁻¹²¹. For example, Gurarie has proposed that strong correlations could realize a topologically non-trivial phase without any surface states when the interactions become sufficiently strong¹¹⁸.

In this chapter, we investigate the question whether electron-electron interactions on the can induce a band gap in the surface electronic band structure. In particular, we focus on the mechanism of **chiral symmetry breaking** which describes the spontaneous generation of an exciton mass of 2D Dirac fermions in the presence of long-range Coulomb interactions above a critical interaction strength. The main idea behind this is to reduce the average kinetic energy of the surface Dirac fermions in order to increase the effective interaction strength, whereas the properties of the bulk should remain unchanged. In the context of 3D topological insulators, this gap creation implies that there are no gapless Dirac-like surface states on the 2D surfaces, although the bulk electronic band structure of the topological insulator remains gapped with a non-trivial band topology. In the following, we investigate under which conditions chiral symmetry breaking could be expected on the surfaces of topological insulators and discuss its possible realization in different models. We show that (i) long-ranged Coulomb interactions cannot drive the transition due to screening, but (ii) short-range Hubbard-like interactions may give rise to gapped surface states, provided the system has an “approximate” chiral symmetry.

This chapter is organized as follows: After an introduction of the mechanism of chiral symmetry breaking in the context of Dirac fermions in two spatial dimensions, we briefly review the

current experimental and theoretical status regarding chiral symmetry breaking in the context of two-dimensional graphene sheets¹²².

We then introduce the two-dimensional **Kane-Mele model** as an effective low-energy model of a topological insulator on the two-dimensional graphene lattice^{3,4}. In particular, we review that the zero-energy edge states in finite-size graphene ribbons with zigzag edges form completely flat bands, but those are not observed for armchair edges. We also show how those flat bands are related to the topology of a whole family of one-parameter Bloch Hamiltonians, and we give explicit criteria for the existence of those zero-energy boundary states for topological insulators in two and three dimensions¹⁸. After that we discuss under which conditions the one-dimensional edge states in graphene ribbons spontaneously develop long-range ferromagnetic order in the presence of Coulomb interactions.

The results on the two-dimensional graphene lattice can be generalized to strong topological insulators described by the **Fu-Kane-Mele model** on the three-dimensional diamond lattice^{7,8}. This model shows both weak and strong topological insulator phases when tuning its model parameters appropriately, but does not exhibit a trivial band insulator phase in the perturbatively accessible regimes. We also consider surfaces of slabs with different orientations described by the Fu-Kane-Mele model, and for certain surfaces we predict the existence of zero-energy states which form completely flat bands. Furthermore, we confirm the existence of those surface bands by explicit tight-binding calculations and discuss the role of spin-orbit interactions generated by second-nearest neighbor hopping. In the presence of Coulomb interactions, the corresponding surface Dirac fermions are expected to show a similar phase diagram as the Dirac fermions in graphene, since the 2D surface state bands become increasingly flat as we decrease the spin-orbit coupling parameter. However, being of 3D nature, we also have to consider the bulk properties of the model, in particular the **dielectric polarization**. We show by an explicit calculation of the polarization function to leading order in the spin-orbit coupling that the topological criterion for the existence of flat surface bands also leads to a sharp peak in the density of states. As a consequence, this activates **screening** of the long-range Coulomb interactions, and chiral symmetry breaking due to Coulomb interactions is avoided. However, due to the vanishing kinetic energy of the Dirac fermions in the flat bands, even local interactions can lead to a spontaneous gap generation. We discuss a possible scenario for this effect in a mean-field approach.

Another class of 3D topological insulators is exemplified by the **minimal model** developed in chapter 3, wherein a similar situation for chiral symmetry breaking is expected to occur. In that system, we also find surface bands which become more and more flat as we approach the bulk critical point. In contrast to the Fu-Kane-Mele model, the origin of the flat surface bands is “spectral pressure” exerted by the bulk conduction and valence bands rather than a topological argument. However, due to the nearby presence of those bulk states, the surface quantum phase transition towards a gapped state is hampered by the diverging polarization function for which we give an explicit calculation. This divergence can also be understood from the presence of an **intermediate semimetallic phase** between the strong topological insulator and the trivial band insulator phases that generically exists in inversion-asymmetric systems such as HgTe, as shown by Murakami^{123,124}.

In the last two parts of this chapter, we discuss the role of short-ranged, Hubbard-like interactions on the surface states which become relevant for systems when long-ranged interactions are screened. In particular, we show that in the Fu-Kane-Mele model the surface states should quite generically develop a surface band gap, whereas a similar situation in the minimal model can only be obtained by fine-tuning the interaction strength. Furthermore, we put our results in a broader context of 3D topological insulators, and discuss the relevance of long-

range Coulomb interactions or local interactions in view of chiral symmetry breaking as a general mechanism for the surface states of topological insulators. This opens up the possibility to identify classes of model Hamiltonians for 3D topological insulators with gapped 2D surface states.

5.1 Coulomb interactions and chiral symmetry breaking

To describe the mechanism of chiral symmetry breaking, we first briefly review the differences between the Fermi liquid description of interacting electrons in a normal metal and the properties of Dirac fermions. Let us start by considering the behavior of electrons in a normal metal, where the electrons interact with each other and with the ions in the crystal via strong long-range Coulomb interactions. Assuming that the electrons propagate freely in a non-relativistic way, the system is Galilean invariant, and the kinetic energy of the electrons is given by the quadratic dispersion $\mathbf{p}^2/(2m)$. Moreover, since electrons carry spin 1/2 and obey Fermi-Dirac statistics, all states below (above) the Fermi energy \mathcal{E}_F are occupied (empty). However, it is not immediately clear that these two basic ideas—Galilean invariance and Fermi-Dirac statistics—are also valid in the presence of strong interactions. Nevertheless, the theory of **screening** due to Lindhard explains that metals are dynamically polarizable materials and that electrons collectively screen the electric fields in the interior of metals, so that the long-range Coulomb interactions become effectively short-ranged and weak enough for a perturbative approach¹²⁵. Altogether, this has led to the theoretical framework of **Landau's Fermi liquid theory**, where the excitations of strongly interacting electrons in a metal can be described in terms of weakly-interacting quasi-particles that carry the same quantum numbers, *e.g.*, spin, as the original particles, but may also have, for example, different effective masses than the bare particles¹²⁶. There are, however, also exceptions to this, where the Fermi liquid ground state becomes unstable. For example, electron-phonon interactions can induce an effective attractive interaction between electrons leading to the Cooper instability of the Fermi liquid towards a superconducting ground state¹²⁷⁻¹²⁹, or the presence of Fermi surface nesting can lead to Fermi surface instabilities resulting in charge or spin density wave ground states¹³⁰⁻¹³². Note that when the electronic density becomes very low and screening becomes ineffective, this also leads to a Fermi liquid instability towards the Wigner crystal state¹³³⁻¹³⁵. The following discussion is due to Kotov *et al.* (Ref. 122).

To develop a better understanding of the interplay between kinetic energy and Coulomb interactions, let us consider electrons moving freely in a metal in a non-relativistic way. Due to the Galilean invariance of such a system, the kinetic energy of a particle is given by $\mathbf{p}^2/(2m)$, where m is the (effective) electron mass, so that the **average kinetic energy** per particle can be estimated as

$$\langle \mathcal{E}_{\text{kin}} \rangle \sim \frac{\hbar^2}{2ml^2} = \frac{\hbar^2}{2m} n_d^{2/d}. \quad (5.1)$$

Here, we have used the fact that the momentum $\mathbf{p} = \hbar\mathbf{k}$ of the particle is of the order $|\mathbf{p}| = \hbar/l$, where the average distance between electrons is given by l , and $n_d \equiv l^{-d}$ denotes the average electron density in d spatial dimensions. On the other hand, the Coulomb interaction takes the usual form

$$V(\mathbf{r}) = \frac{e^2}{\epsilon |\mathbf{r}|}, \quad (5.2)$$

where e is the electron charge, and ϵ denotes the dielectric constant of the host medium. With the average distance $|\mathbf{r}| = l$ between two particles, the **average Coulomb interaction energy**

is of the order

$$\langle \mathcal{E}_{\text{Coul}} \rangle \sim \frac{e^2}{\epsilon} n_d^{1/d}, \quad (5.3)$$

thus following a different power-law behavior with respect to the electron density. The ratio of the average Coulomb energy $\langle \mathcal{E}_{\text{Coul}} \rangle$ to the average kinetic energy per particle $\langle \mathcal{E}_{\text{kin}} \rangle$ is usually referred to as **interaction parameter** r_S which is given by

$$r_S \equiv \frac{\langle \mathcal{E}_{\text{Coul}} \rangle}{\langle \mathcal{E}_{\text{kin}} \rangle} = \frac{2me^2}{\epsilon \hbar^2} n_d^{-1/d} = \left(\frac{n_0}{n_d} \right)^{1/d}, \quad (5.4)$$

where, for brevity, we have introduced the electron density n_0 which depends on the material properties as follows:

$$n_0 \equiv \left(\frac{2me^2}{\epsilon \hbar^2} \right)^d. \quad (5.5)$$

If the average electron density n_d is relatively high, *i.e.*, $n_d \gg n_0$, then the kinetic energy dominates over the Coulomb energy, and the latter can be considered as a perturbation to the free electron gas. In that limit, the Fermi liquid description of electrons in a metal is well-defined and yields the correct results. However, upon lowering the average electron density such that $n_d \ll n_0$, the Coulomb energy becomes more and more dominant and can lead to instabilities of the free electron gas like Wigner crystallization^{133–135}. Therefore, in a Galilean invariant system, the relative strength r_S of the Coulomb interaction energy and the kinetic energy is determined by the electron density which, to a certain extent, can be controlled experimentally by changing the electron density, opening up the possibility to study different electronic phases aside from the Fermi liquid phase.

Although a large number of theoretical works on the two-dimensional honeycomb lattice existed, showing that single-layer graphene sheets have interesting properties, the experimental discovery of **graphene** in 2004¹³⁶ has proven that the picture presented above has to be extended. Since the carbon atoms in graphene are organized in a regular hexagonal lattice in two dimensions, the low-energy properties of the electrons close to the Fermi level are described by the linear Dirac dispersion $\pm v_F |\mathbf{p}|$, where v_F is the Fermi velocity. Although v_F is a material property and much smaller than the speed of light c , *i.e.*, $v_F/c \approx 1/300$ for graphene, the effective low-energy theory of electrons in graphene is much more similar to the one of 2D Dirac fermions, and thus quite distinct from the usual Galilean invariant formulation of Landau's Fermi liquid theory. For example, we immediately recognize that the average kinetic energy per electron in two-dimensional graphene is of the order

$$\langle \mathcal{E}_{\text{kin}} \rangle \sim \frac{\hbar v_F}{l} = \hbar v_F n_d^{1/2} \quad (5.6)$$

and that it shows a different behavior compared to the Fermi liquid case in $d = 2$, where $\langle \mathcal{E}_{\text{kin}} \rangle \sim n_d$ [cf. Eq. (5.1)]. Furthermore, due to the linear dispersion the electronic density of states vanishes linearly at the Dirac point, leading to the notion of graphene being a **semimetal**, *i.e.*, a zero-band gap semiconductor. As a consequence, in contrast to the electron gas described by Fermi liquid theory, pristine graphene cannot screen the long-range Coulomb interaction due to the lack of electronic states at the Fermi level. To compare the average kinetic energy with the Coulomb interaction, note that the electron-electron interactions can be considered as instantaneous, because the photons mediating the Coulomb interaction are much faster than the electrons in graphene, $v_F \ll c$. Therefore, the Coulomb interaction in graphene takes the

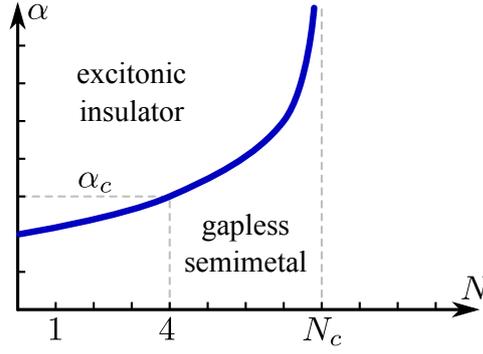


Figure 5.1. Schematic zero-temperature phase diagram of chiral symmetry breaking in QED_{2+1} as function of the number of fermion flavors N and the interaction strength $\alpha = e^2/(\hbar v_F)$. A dynamical mass $\Delta(\epsilon, \mathbf{p})$ is generated if N is below a critical number $N_c \approx 7 - 9$ and is found above a critical interaction strength $\alpha_c \approx 0.8 - 3.3$ (Refs. 15,138–141). The above numerical results are based on various approximations in different limits. For graphene with physical flavor number $N = 4$ (valley and spin degrees of freedom) a similar phase transition towards the gapped exciton insulator is expected once a critical interaction strength α_c has been reached. Furthermore, in the simplest case of a strong topological insulator, there is a single Dirac cone located at the surfaces, *i.e.*, we have $N = 1$. Figure after Ref. 122.

usual form (5.2) which, however, breaks the Lorentz invariance of the low-energy theory. Nevertheless, if we compare the kinetic energy and the Coulomb interaction energy we observe that this ratio

$$\alpha \equiv \frac{\langle \mathcal{E}_{\text{Coul}} \rangle}{\langle \mathcal{E}_{\text{kin}} \rangle} = \frac{e^2}{\hbar \epsilon v_F} \quad (5.7)$$

is actually independent of the electron density n_d , but depends on (i) material properties such as the Fermi velocity v_F and (ii) environmental properties like the dielectric constant ϵ . In the case of weak electron-electron interactions, *i.e.*, for $\alpha \ll 1$, we can use standard perturbative methods to calculate, for example, the fermion self-energy $\Sigma(\nu, \mathbf{k})$ or the polarization function $\Pi(\omega, \mathbf{q})$. In particular, the weak-coupling analysis shows that $\alpha = 0$ is an infrared-stable fixed point, and thus Coulomb interactions are marginally irrelevant up to logarithmic corrections^{122,137}. Nevertheless, in the strong coupling limit, long-range Coulomb interactions can also lead to spontaneous symmetry breaking of the ground state due to an excitonic pairing mechanism which is discussed below.

In relativistic quantum electrodynamics in $2 + 1$ dimensions or QED_{2+1} , Dirac fermions in the presence of long-range Coulomb interactions can exhibit a spontaneously generated mass term, thereby creating a gapped system. This phenomenon is known as **chiral symmetry breaking** in the literature and is subject of current research^{142–147}, although the idea actually goes back more than 25 years^{148–150}. The analysis in relativistic QED_{2+1} shows that a dynamical mass $\Delta(\epsilon, \mathbf{p})$ can be generated if the number of fermion flavors N is below a critical number N_c . Moreover, for a fixed value of N , a transition to a gapped state is only found above a critical interaction strength $\alpha_c(N)$. The overall phase diagram at zero temperature as function of the number of fermion flavors N and the interaction strength α is expected to look similar to the one shown in Fig. 5.1. We also briefly note that at finite temperatures one expects the existence of a critical temperature given by $k_B T_c \sim \Delta(0)$, while finite doping μ is expected to quickly destroy the dynamical gap¹³⁹.

However, there exist important differences between QED_{2+1} and graphene: First, note that the Dirac fermions are confined to the 2D sheet of graphene, whereas the field lines of the electric fields responsible for the Coulomb interaction extend through 3D space. Second, since the

Fermi velocity v_F of the Dirac fermions in graphene is much smaller than the speed of light the Coulomb interaction can be considered as instantaneous. As a consequence, the Coulomb interaction breaks the Lorentz invariance, showing that the Dirac-like quasi-particles arise only due to the peculiar band structure of graphene, and in that case the mass scale is set by the coupling itself. In graphene, there exists also the possibility of a dynamically generated mass, but there the mass scale is related to the ultraviolet momentum cutoff Λ , where the bandwidth of the Dirac fermions is given by $W = v_F \Lambda$. Furthermore, in graphene the number of Dirac fermions is fixed by $N = 4$ taking the spin and valley degrees of freedom into account. On the other hand, in strong topological insulators the fermion flavor number is equal to $N = 1$ in the simplest case, where the surfaces host a single Dirac cone.

The dynamic mass $\Delta(\epsilon, \mathbf{p})$ can be obtained as a self-consistent solution within the random-phase approximation (RPA) for the electronic self-energy $\Sigma(\nu, \mathbf{k})$ from the following **Dyson-Schwinger equation**¹²²:

$$\Delta(\epsilon, \mathbf{p}) = i \int_{-\infty}^{\infty} \frac{d\nu}{2\pi} \int \frac{d^2k}{(2\pi)^2} \frac{V_{\text{RPA}}^{2\text{D}}(\nu - \epsilon, \mathbf{k} - \mathbf{p}) \Delta(\nu, \mathbf{k})}{\nu^2 - v_F^2 |\mathbf{k}|^2 - \Delta(\nu, \mathbf{k})^2 + i0^+}. \quad (5.8)$$

This mass gap $\Delta(\epsilon, \mathbf{p})$ has a strong momentum dependence due to the long-range nature of the Coulomb interaction, with $\Delta(\epsilon, \mathbf{p})$ reaching its maximal value at small momenta and decreasing at large momenta¹³⁸. The dynamical Dyson-Schwinger equation has been studied both analytically and numerically using various approximation schemes^{138,139}, and a number of predictions for the critical fermion flavor number N_c and the critical interaction strength α_c have been made. For example, a significant simplification of the self-consistency equation is obtained in the static limit, where the potential $V_{\text{RPA}}(\omega, \mathbf{q})$ is replaced by a static one, $V_{\text{RPA}}(\omega = 0, \mathbf{q})$ (Ref. 15). However, since all of the methods are based on various approximations and different limits, the results vary in a certain range, but it is generally accepted that α_c lies within 0.8 – 3.3, while $N_c \approx 7 - 9$ (Refs. 15,138–141).

The physical structure of the gapped state depends on the order parameter for the chiral symmetry breaking. In the case of a two-component Dirac theory described by the following Dirac Hamiltonian,

$$\mathcal{H}(\mathbf{k}) = v_F \mathbf{k} \cdot \boldsymbol{\sigma} + m_z \sigma_z = v_F \begin{bmatrix} m_z & k_x - ik_y \\ k_x + ik_y & -m_z \end{bmatrix}, \quad (5.9)$$

a gap in the spectrum can be opened, for example, by a mass term $m_z \sigma_z$ coupling to the third Pauli matrix, because mass terms like $m_x \sigma_x$ or $m_y \sigma_y$ can be treated by a suitable redefinition of the two-dimensional surface momentum $\mathbf{k} = (k_x, k_y)^T$:

$$\mathcal{E}_{\pm}(\mathbf{k}) = \pm \sqrt{v_F^2 |\mathbf{k}|^2 + m_z^2}. \quad (5.10)$$

Consequently, to describe chiral symmetry breaking, one investigates the spontaneous mass generation in terms of the order parameter $\langle \Psi^\dagger \sigma_z \Psi \rangle$. In graphene, many different symmetry-broken states can occur due to chiral symmetry breaking. For example, in graphene, σ_z may correspond to the sublattice degree of freedom, and thus the gapped state is a charge density wave state with a modulation of the electronic density breaking the sublattice inversion symmetry. On the other hand, on the surfaces of strong topological insulators the spinor degrees of freedom are associated with the electron spin due to the helical nature of the surface state, implying that the chiral symmetry breaking corresponds to spontaneous ferromagnetism which opens a spin gap and breaks time-reversal symmetry on the surface. Note that once such a gap

is opened one expects to see, for example, the anomalous quantum Hall effect or the surface magneto-electric effect associated with topological insulators^{11–13}.

Finally, concerning the interaction parameter $\alpha = e^2/(\hbar\epsilon v_F)$ [cf. Eq. (5.7)] we observe that there are, in principle, two ways to increase the interaction strength α by (i) decreasing the dielectric constant ϵ or (ii) reducing the Fermi velocity v_F of the Dirac fermions. The first strategy has been pursued extensively in the context of graphene. Here, by suspending the graphene sheet one can control the dielectric constant ϵ through the environment, and one expects a value of $\alpha \approx 2.2$ in vacuum¹⁷. Although this is well within the range of the reported critical values $\alpha_c \approx 0.9 - 3.3$, there is no experimental evidence which would confirm the existence of an excitonic insulator phase in graphene. Instead, recent quantum oscillation measurements show a semimetallic behavior down to lowest fillings and temperatures¹⁵¹. The apparent contradiction between theory and experiment is subject of current research.

In the context of topological insulators, our basic idea is to pursue the second strategy, *i.e.*, we would like to decrease the Fermi velocity v_F of the surface Dirac fermions. In contrast to graphene, where the effective interaction strength $\alpha = e^2/(\hbar\epsilon v_F)$ is bound from above as ϵ is bound from below by its vacuum value, this approach in principle allows to drastically increase α and thereby opens up the possibility for a surface quantum phase towards a gapped state. We will investigate whether this can be achieved by approaching the bulk quantum critical point of the topological insulator, where the bulk band gap closes and a phase transition to a topologically trivial phase occurs. The appropriate knob to control the Fermi velocity is the spin-orbit coupling parameter which has been demonstrated experimentally via chemical substitution in bismuth-based compounds⁷², but can in principle also be achieved in mercury telluride systems by replacing mercury atoms with cadmium atoms⁶. In principle, this tuning of spin-orbit coupling works, provided the dielectric constant ϵ characterizing the bulk of the sample remains finite, but we will have to investigate as well to what extent ϵ is affected by the adiabatic change in the spin-orbit coupling.

To develop an explicit understanding of the spontaneous mass generation on the surfaces of topological insulators, we first concentrate on a toy model which was introduced by Kane and Mele on the graphene lattice^{3,4} and later generalized to the three-dimensional diamond lattice by Fu, Kane, and Mele^{7,8}. In those models, we show that the surface states form completely flat bands when the spin-orbit interaction parameter approaches a critical value of zero. Notably, at the quantum critical point the models possess an additional chiral symmetry which guarantees the existence of zero-energy states at the boundaries of the system. While in graphene screening of long-range Coulomb interactions is absent due to the lack of states at the Fermi level and thus a spontaneous mass generation is expected to occur, the situation turns out to be completely different on the 3D diamond lattice. Here, the nodal lines with zero energy in the bulk of the system, which enforce the existence of the zero-energy surface states, also lead to screening of the Coulomb interactions, and therefore the effective interaction strength α will not diverge. Nevertheless, since the average kinetic energy of the surface states is zero, even local interactions can lead to spontaneous gap generation.

After that we will consider another class of topological insulators which is exemplified by the minimal model based on mercury telluride, as discussed in chapter 3. Here, we find a specific situation for a potential enhancement of the long-range Coulomb interactions by means of decreasing the Fermi velocity v_F of the surface Dirac fermions without any additional fine-tuning and symmetries. Unfortunately, the same effect which renders v_F small also increases the dielectric constant ϵ and finally prevents the interaction strength α from becoming large.

We also discuss the effect of local interactions on the surface electronic band structure and show that local interactions can generically lead to a surface band gap if the surface states form-

ing the flat band do not penetrate too much into the bulk. We find such a situation in the Fu-Kane-Mele model, while the minimal model again evades a spontaneous gap generation.

5.2 The Kane-Mele model of 2D topological insulators

The **Kane-Mele Hamiltonian** is one of the most well-studied toy models in the context of 2D topological insulators and it was introduced in 2005 by Kane and Mele^{3,4}. It is inspired by the Haldane model, a mathematical model for spinless fermions under the influence of a fictitious inhomogeneous magnetic field with vanishing total flux (see chapter 2). Although time-reversal symmetry is broken in the Haldane model, it opens up the possibility to observe the quantum Hall effect in the absence of an applied external magnetic field, as discussed in chapter 2. By including the spin degree of freedom and by making the inhomogeneous magnetic field spin-dependent, Kane and Mele have restored the time-reversal symmetry of the system, giving rise to the **quantum spin Hall insulator** phase. In the original work, Kane and Mele have studied graphene as a 2D topological insulator and, in particular, the competition between the intrinsic spin-orbit interaction and inversion-symmetry breaking perturbations like a Rashba spin-orbit interaction and a staggered perturbation, which make graphene an ordinary insulator^{3,4}. In the following, we discuss the basic aspects of the graphene lattice and the spin-orbit interaction in graphene, before we formulate the Kane-Mele Hamiltonian and discuss its properties and its phase diagram. After that, we consider the appearance of edge states in graphene ribbons and discuss their properties as a function of various model parameters. Furthermore, following Ref.¹⁸ we describe a topological argument for the existence of zero-energy edge states. We show that those edge states can become perfectly flat which opens up the possibility for spontaneous ferromagnetism at the edges of graphene ribbons.

5.2.1 The crystal structure of the graphene lattice

Graphene is the ideal 2D form of carbon, where the atoms are arranged in a hexagonal lattice or **honeycomb lattice** structure, as shown in Fig. 5.2 (a). This planar structure of carbon atoms is characterized by two types of bonds which exhibit the so-called **sp^2 hybridization**. Of the four valence orbitals the $2s$, $2p_x$, and $2p_y$ orbitals form the in-plane bonding and anti-bonding σ , σ^* orbitals by superpositions, and the σ orbitals of neighboring carbon atoms form strong covalent bonds which are responsible for the elastic properties of the 2D graphene sheet. Moreover, note that the σ , σ^* orbitals are even with respect to the planar symmetry. The remaining $2p_z$ orbital, however, points out of the plane and does not couple to the σ states, because it is odd with respect to the planar mirror symmetry of the graphene sheet. The interaction between neighboring $2p_z$ orbitals then creates the delocalized bonding and anti-bonding π , π^* orbitals which are responsible for the electronic properties of graphene.

The honeycomb lattice actually is a triangular lattice with a basis of two atoms per unit cell, where the **primitive lattice vectors** are given by

$$\mathbf{a}_1 = \frac{a}{2}(3, \sqrt{3})^T, \quad \mathbf{a}_2 = \frac{a}{2}(3, -\sqrt{3})^T, \quad (5.11)$$

and a denotes the distance between neighboring atoms [cf. Fig. 5.2 (a)]. The position of each atom in the graphene lattice can be decomposed into $\mathbf{r}_{jl} = \mathbf{R}_j + \mathbf{r}_l$, where \mathbf{R}_j denotes the position of the j th unit cell of the Bravais lattice, and \mathbf{r}_l is the position of the l th atom within the unit cell. The two atoms in the unit cell of the honeycomb lattice are located at $\mathbf{r}_1 = (0, 0)^T$ and $\mathbf{r}_2 = \frac{a}{2}(1, \sqrt{3})^T$. Hence, the positions of the three nearest neighbors of a carbon atom are

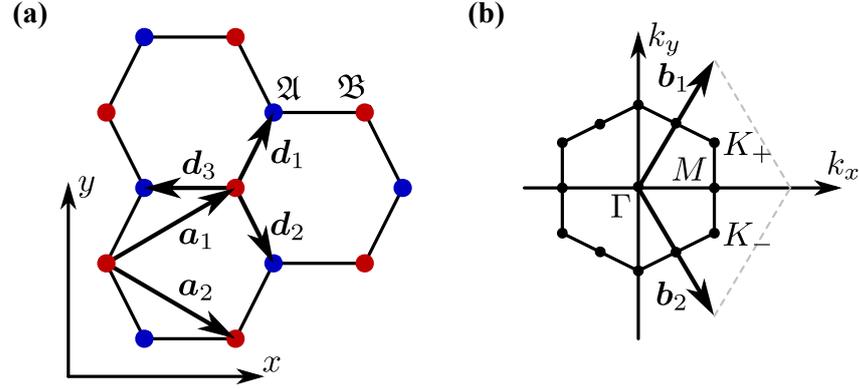


Figure 5.2. (a) The real-space lattice structure of graphene consisting of two interpenetrating triangular lattices forming two sublattices \mathfrak{A} and \mathfrak{B} (indicated by blue and red dots). \mathbf{a}_1 and \mathbf{a}_2 are the primitive lattice vectors, and \mathbf{d}_i with $i = 1, 2, 3$ denote the bond vectors to nearest-neighboring atoms. (b) The reciprocal lattice of graphene and its first Brillouin zone. Points of high-symmetry are given by $\Gamma = (0, 0)^T$ and $M = \frac{2\pi}{3a}(1, 0)^T$, and the Dirac cones are located at the two non-equivalent corners of the Brillouin zone, $\mathbf{K}_{\pm} = \frac{2\pi}{3a}(1, \pm 1/\sqrt{3})^T$.

given by

$$\mathbf{d}_1 = \frac{a}{2}(1, \sqrt{3})^T, \quad \mathbf{d}_2 = \frac{a}{2}(1, -\sqrt{3})^T, \quad \mathbf{d}_3 = a(-1, 0)^T, \quad (5.12)$$

while the six second-nearest neighbors are located at positions

$$\mathbf{d}'_{1\pm} = \pm\mathbf{a}_1, \quad \mathbf{d}'_{2\pm} = \pm\mathbf{a}_2, \quad \mathbf{d}'_{3\pm} = \pm(\mathbf{a}_2 - \mathbf{a}_1). \quad (5.13)$$

The **reciprocal lattice** of the graphene lattice is given by two **primitive reciprocal lattice vectors** \mathbf{b}_1 and \mathbf{b}_2 which are defined by the relation $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$, where δ_{ij} is the Kronecker δ -symbol. The reciprocal lattice points of the triangular lattice form again a triangular lattice which is spanned by the primitive reciprocal lattice vectors

$$\mathbf{b}_1 = \frac{2\pi}{3a}(1, \sqrt{3})^T, \quad \mathbf{b}_2 = \frac{2\pi}{3a}(1, -\sqrt{3})^T, \quad (5.14)$$

as shown in Fig. 5.2 (b). The first **Brillouin zone** of the honeycomb lattice also forms a regular hexagon, but rotated by 90° in the xy plane with respect to the real-space honeycomb lattice. Note that the symmetries of the Brillouin zone are determined by the symmetries of the crystal lattice due to the construction of the reciprocal lattice vectors \mathbf{b}_j from the primitive lattice vectors \mathbf{a}_i .

The 2D space group or **plane group** of the hexagonal lattice is symmorphic, indexed as number 17, and denoted by C_{6v}^1 in accord with the Schoenflies notation or $p6mm$ in international notation^{152,153}. Here, the symbol $p6mm$ describes a primitive lattice in two spatial dimensions with a six-fold rotation axis normal to the lattice plane and reflections across the $\{10\}$ and the $\{21\}$ mirror lines, respectively. On the other hand, the symmetries at the center of the Brillouin zone, *i.e.*, the Γ point, are described by the hexagonal **point group** C_{6v} . The 12 elements of this point group can be divided into 6 classes $\{E\}$, $\{C_2\}$, $\{2C_3\}$, $\{2C_6\}$, $\{3\sigma_d\}$, $\{3\sigma_v\}$ and are listed in Table 5.1. Here, note the presence of vertical reflection planes σ_v and diagonal reflection planes σ_d which, for clarity, are illustrated in Fig. 5.3. The point group C_{6v} has six **irreducible representations** which are usually denoted by $A_1, A_2, B_1, B_2, E_1,$ and E_2 . Note that in condensed matter physics the wave functions are usually classified by their symmetries at the Γ point, while the above terminology of irreducible representations is more commonly used in molecular physics.

class	symmetries of the plane group #17 / $C_{6v} / p6mm$
$\{E\}$	identity
$\{C_2\}$	two-fold rotation of 180° about the $[001]$ axis
$\{2C_3\}$	three-fold clockwise and counterclockwise rotations of 120° about the $[001]$ axis
$\{2C_6\}$	six-fold clockwise and counterclockwise rotations of 60° about the $[001]$ axis
$\{3\sigma_v\}$	three reflections in a vertical plane containing the axis of highest rotational symmetry (z axis) and one of the primitive lattice vectors
$\{3\sigma_d\}$	three reflections in a diagonal plane containing the axis of highest rotational symmetry and bisecting the angle between two σ_v reflections

Table 5.1. The 12 elements of the plane group C_{6v}^1 and the corresponding point group C_{6v} are divided into the above six symmetry classes, leading to six distinct irreducible representations.

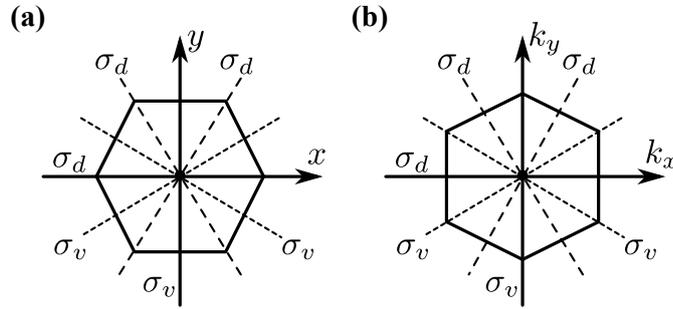


Figure 5.3. Illustration of the reflection symmetries σ_v and σ_d in a 2D graphene sheet in (a) real space and (b) reciprocal space.

5.2.2 Tight-binding approach to the electronic band structure of graphene

Although carbon has four valence orbitals, the $2s$, $2p_x$, and $2p_y$ orbitals hybridize forming the in-plane bonding and anti-bonding σ , σ^* orbitals, while the perpendicular $2p_z$ orbitals form the delocalized bonding and anti-bonding π , π^* orbitals which are mainly responsible for the electronic properties of graphene. Taking the electron's spin into account, we have to consider eight states per carbon atom, *i.e.*, a total of 16 states in the unit cell of the honeycomb lattice. In principle, the resulting 16×16 Hamiltonian can be described by a tight-binding Hamiltonian using the Slater-Koster parametrization of the overlap parameters similar to the mercury telluride system discussed in chapter 3. Note that the bonding and anti-bonding σ , σ^* and π , π^* bands have different parity with respect to in-plane mirror symmetry of the graphene sheet, and therefore matrix elements coupling those bands vanish exactly. Moreover, since the σ , σ^* bands are well separated in energy from the π , π^* bands and also too far away from the Fermi level to play an important role for the electronic properties of graphene, they are usually neglected in the tight-binding approach. As a consequence, we focus on the two π , π^* bands and derive an effective 4×4 Hamiltonian describing the electronic band structure of graphene which gives rise to the electronic valence and conduction bands linearly crossing the Fermi level at the high-symmetry points \mathbf{K}_\pm in the Brillouin zone of graphene.

Let us start by introducing two sublattices \mathfrak{A} and \mathfrak{B} for the graphene lattice with annihilation operators a , a^\dagger and b , b^\dagger acting on the corresponding \mathfrak{A} and \mathfrak{B} sublattices, respectively. Let us define a four-component spinor $\Psi_i = (a_{i\uparrow}, a_{i\downarrow}, b_{i\uparrow}, b_{i\downarrow})^T$ describing the sublattice and spin degrees of freedom, where i labels the crystallographic unit cells. Within the tight-binding ap-

proximation, the electronic interaction between neighboring carbon atoms is then described by the following effective Hamiltonian:

$$H_0 = -t \sum_{\langle i,j \rangle, \sigma} (c_{i\sigma}^\dagger c_{j\sigma} + \text{H.c.}) = -t \sum_{\langle i,j \rangle, \sigma} (a_{i\sigma}^\dagger b_{j\sigma} + \text{H.c.}), \quad (5.15)$$

where the usual fermionic creation and annihilation operators c, c^\dagger act on both \mathfrak{A} and \mathfrak{B} sublattice. The energy bands of that Hamiltonian can be easily calculated by diagonalizing the corresponding Bloch Hamiltonian. First, note that the Fourier transform of the four-component spinor Ψ_i is given by

$$\Psi_i = \frac{1}{\sqrt{N}} \sum_{\mathbf{k}} e^{i\mathbf{k} \cdot \mathbf{R}_i} \Psi_{\mathbf{k}}, \quad \Psi_{\mathbf{k}} = \frac{1}{\sqrt{N}} \sum_i e^{-i\mathbf{k} \cdot \mathbf{R}_i} \Psi_i, \quad (5.16)$$

where N denotes the number of unit cells in the crystal, and \mathbf{R}_i denotes the position of the i th unit cell. Upon substituting the above expression for Ψ_i it is straightforward to show that the corresponding Bloch Hamiltonian $\mathcal{H}_0(\mathbf{k})$ is given by the following 4×4 Hermitian matrix:

$$\mathcal{H}_0(\mathbf{k}) = -(t/2)(\gamma(\mathbf{k}) \tau_- \otimes \sigma_0 + \text{H.c.}). \quad (5.17)$$

Here, for brevity, we have introduced two vectors of Pauli matrices, $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)^T$ and $\boldsymbol{\tau} = (\tau_x, \tau_y, \tau_z)^T$, acting on the spin and sublattices degrees of freedom, respectively, and $\sigma_0 = \tau_0 = \mathbb{1}$ is the 2×2 identity matrix. $\tau_{\pm} = \tau_x \pm i\tau_y$ are the usual raising and lowering operators describing hopping between the \mathfrak{A} and \mathfrak{B} sublattice. Finally, $\gamma(\mathbf{k})$ is a sum over phase factors $\exp(i\mathbf{k} \cdot \mathbf{d}_j)$ describing the hopping between nearest-neighbor carbon atoms located at \mathbf{d}_j :

$$\gamma(\mathbf{k}) \equiv \sum_{j=1}^3 e^{i\mathbf{k} \cdot \mathbf{d}_j}. \quad (5.18)$$

To compute the electronic band structure of graphene, we diagonalize the Bloch Hamiltonian $\mathcal{H}(\mathbf{k})$, and due to the fact that the matrices $\tau_- \otimes \sigma_0$ and $\tau_+ \otimes \sigma_0$ anti-commute we obtain the following eigenvalues of $\mathcal{H}_0(\mathbf{k})$:

$$\mathcal{E}_{\pm}(\mathbf{k}) = \pm t |\gamma(\mathbf{k})|, \quad (5.19)$$

where the \pm sign corresponds to the conduction (valence) band. On the honeycomb lattice, one can also compute analytical expressions for the eigenvalues of the Bloch Hamiltonian:

$$\mathcal{E}_{\pm}(\mathbf{k}) = \pm t \sqrt{3 + 2 \cos(\sqrt{3}ak_y) + 4 \cos(3ak_x/2) \cos(\sqrt{3}ak_y/2)}. \quad (5.20)$$

Note that this result was obtained in the absence of spin-orbit interactions and next-nearest neighbour hopping, whose effects on the electronic band structure will be discussed below.

The hexagonal Brillouin zone of graphene has three kinds of high-symmetry points denoted by Γ , K , and M , as shown in Fig. 5.2 (b). First, the direct band gap between the conduction and valence band is largest at the center of the Brillouin zone, *i.e.*, the Γ point. Second, at the M point, the density of states exhibits a van Hove singularity and vanishes logarithmically, because the conduction (valence) bands show a saddle point as function of the crystal momentum \mathbf{k} . Finally, the conduction and valence bands touch at the corners \mathbf{K}_{\pm} of the Brillouin zone, and importantly the density of states vanishes linearly (see Figs. 5.4 and 5.5). Note that there are two non-equivalent points \mathbf{K}_{\pm} in the Brillouin zone of the graphene lattice, which are not connected by a reciprocal lattice vector, and those two points are often referred to as the

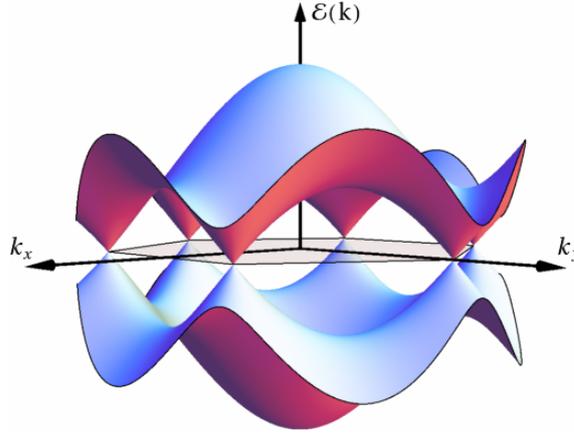


Figure 5.4. Energy spectrum of a single two-dimensional graphene sheet as function of the crystal momentum \mathbf{k} . The first Brillouin zone is indicated by a regular hexagon (opaque shading) with the two non-equivalent \mathbf{K}_{\pm} at the corners the Brillouin zone, where the conduction and valence bands touch linearly as $\mathcal{E}_{\pm}(\mathbf{k}) = \pm v_F |\mathbf{k}|$.

two **valleys** of graphene. Moreover, close to the \mathbf{K}_{\pm} , the conduction and valence bands show a conical behavior which can be seen by substituting $\mathbf{k} = \mathbf{K}_{\pm} + \mathbf{q}$ and expanding the Bloch Hamiltonian for small \mathbf{q} :

$$H_{\pm}(\mathbf{q}) = \pm v_F \mathbf{q} \cdot \boldsymbol{\sigma}. \quad (5.21)$$

Here, $v_F \equiv 3ta/2 \approx 10^6$ m/s is the Fermi velocity in graphene, \mathbf{q} denotes the momentum relative to \mathbf{K}_{\pm} as before, and $\boldsymbol{\sigma} = (\sigma_x, \sigma_y)^T$ is a two-dimensional vector of Pauli matrices. Furthermore, in contrast to our conventions $\boldsymbol{\sigma}$ does not act on the real electronic spin of the electron, but on the sublattice degree of freedom. Hence, σ_z is often referred to as **pseudospin** degree of freedom. The resulting low-energy spectrum (5.22) is formally equivalent to the one obtained from solving the massless Dirac equation in two spatial dimensions:

$$\mathcal{E}_{\pm}(\mathbf{q}) \approx \pm v_F |\mathbf{q}| + \mathcal{O}(q^2). \quad (5.22)$$

Note that this result was obtained by Wallace already in 1947¹⁵⁴. Due to the sp^2 hybridization of each carbon atom, there is only one electron left in the π orbital, and thus the system is half-filled. As a consequence, the transport properties and low-energy physics of graphene are entirely dominated by the conical spectrum (5.22) close to the so-called **Dirac points** \mathbf{K}_{\pm} . For further reading we refer the reader to recent reviews and references therein^{155–158}.

5.2.3 Spin-orbit interaction in graphene

Let us now consider the role of spin-orbit interactions in graphene. As discussed previously, the electronic properties of graphene are determined by the π, π^* orbitals formed by the $2p_z$ orbitals which have orbital angular momentum $l = 1$ and magnetic quantum number $m_l = 0$. The spin-orbit interaction $H_{SO} = \lambda_{SO} \mathbf{l} \cdot \mathbf{s}$, however, couples orbital angular momentum states with different magnetic quantum numbers. For example, spin-up electrons in the $2p_z$ orbitals are coupled to spin-down electrons in the $2p_x$ and $2p_y$ orbitals (see chapter 3 for a more detailed discussion of the spin-orbit interaction between p orbital states). Therefore, spin-orbit interactions couple the electronic π, π^* states with the σ, σ^* states. One approach to include spin-orbit interactions into the tight-binding approach is to give up the effective low-energy description of

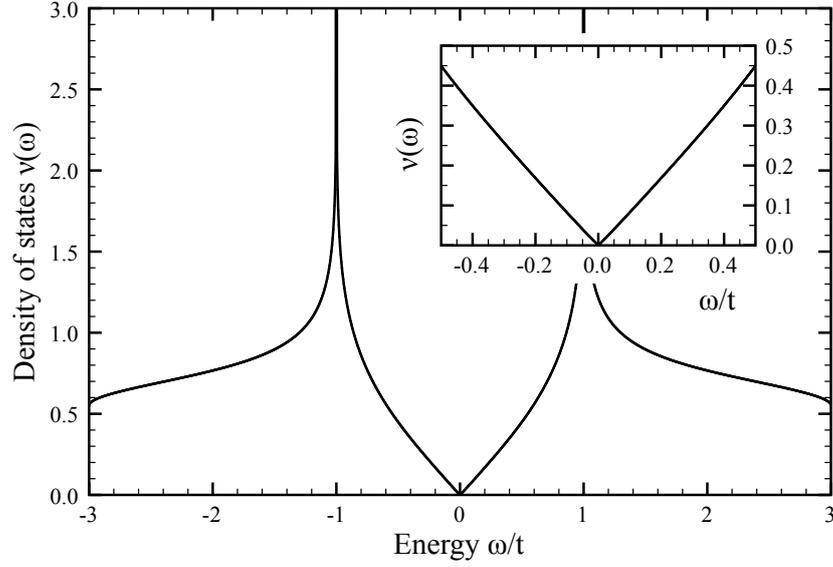


Figure 5.5. Density of states $\nu(\omega)$ per unit cell as a function of energy ω/t computed from the electronic band structure of graphene. $\nu(\omega)$ is symmetric with respect to $\omega = 0$ in the absence of next-nearest neighbor hopping. The inset shows a zoom-in close to the charge-neutrality point of one electron per site, where the density of states is approximately linear in energy, *i.e.*, $\nu(\omega) \propto |\omega|$.

graphene in terms of the π, π^* orbitals only, *i.e.*, to use a tight-binding model Hamiltonian consisting of four atomic orbitals $\{s, p_x, p_y, p_z\}$ and two spin states $\{\uparrow, \downarrow\}$. In that case, the numerical calculation of the electronic band structure shows that the effect of spin-orbit interactions on the overall energy scale of the band structure is not very pronounced, with the exception that a band gap is opened at the Dirac points. However, for realistic values of the spin-orbit coupling in graphene, *i.e.*, $\lambda_{\text{SO}} \approx 10$ meV, this bulk band gap is only of the order 10 μeV (Refs. 3,4,159). Moreover, the numerical calculations and analytical symmetry arguments show that the electron's spin is a good quantum number close to the Dirac points⁴. These results suggest that one can derive an effective Hamiltonian for the π, π^* bands of graphene by integrating out the higher energy bands, neglecting the spin-orbit interactions between the σ and π orbitals. As shown by Kane and Mele, the remaining spin-orbit interaction between spin-up and spin-down states within the π, π^* orbitals can also be described in terms of an effective spin-orbit Hamiltonian, as discussed below.

5.2.4 The Kane-Mele Hamiltonian

An explicit Hamiltonian describing the electronic properties of graphene in the presence of spin-orbit interactions was given by Kane and Mele in 2005^{3,4}, introducing the following effective 4×4 **Kane-Mele Hamiltonian** for graphene:

$$\begin{aligned}
 H_{\text{KM}} = & -t \sum_{\langle i,j \rangle, \sigma} (c_{i\sigma}^\dagger c_{j\sigma} + \text{H.c.}) + i\lambda_{\text{SO}} \sum_{\langle\langle i,j \rangle\rangle, \sigma, \sigma'} c_{i\sigma}^\dagger (\boldsymbol{\nu}_{ij} \cdot \boldsymbol{\sigma})_{\sigma\sigma'} c_{j\sigma'} \\
 & + i\lambda_{\text{R}} \sum_{\langle i,j \rangle, \sigma, \sigma'} c_{i\sigma}^\dagger [(\boldsymbol{\sigma} \times \hat{\mathbf{d}}_{ij}) \cdot \mathbf{e}_z]_{\sigma\sigma'} c_{j\sigma'} + \Delta \sum_{i, \sigma} \xi_i c_{i\sigma}^\dagger c_{i\sigma},
 \end{aligned} \tag{5.23}$$

where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)^T$ acts on the electron's spin degree of freedom. The first term represents the usual matrix elements describing the hopping of electrons between nearest-neigh-

boring atoms, where t denotes the spin-independent **overlap parameter** or **hopping amplitude**. The second term describes the **spin-orbit interaction** of the conduction electrons. Since the Kane-Mele Hamiltonian is formulated in terms of electrons with s orbital symmetry, the spin-orbit interaction is effectively generated by an interaction which connects second-nearest neighbors with a spin-dependent amplitude $\pm\lambda_{\text{SO}}$. Note that on the graphene lattice the vector $\boldsymbol{\nu}_{ij}$ is defined by

$$\boldsymbol{\nu}_{ij} \equiv (0, 0, \nu_{ij}^z)^T \equiv \frac{1}{\mathcal{N}} \left[\mathbf{d}_{ij}^{(1)} \times \mathbf{d}_{ij}^{(2)} \right], \quad (5.24)$$

where $\mathbf{d}_{ij}^{(1,2)}$ denote the two bond vectors which are traversed by the electron when moving from site i to site j . The normalization prefactor $\mathcal{N} = \sqrt{3}a^2/2$ is chosen such that $\boldsymbol{\nu}_{ij}$ is a unit vector, *i.e.*, $\nu_{ij}^z = \pm 1$. The third term in Eq.(5.23) is a **Rashba spin-orbit interaction** with strength λ_{R} between electrons located on nearest neighboring atoms connected by the bond vector \mathbf{d}_{ij} . Such a term, which explicitly violates the mirror symmetry of the graphene sheet, may be generated, for example, by a perpendicular electric field or by interactions with the substrate. Finally, the last term describes a **staggered sublattice potential** with $\xi_i = \pm 1$ on the \mathfrak{A} and \mathfrak{B} sublattice, respectively, which breaks the two-fold rotation symmetry of the graphene lattice. This term is included in the Kane-Mele Hamiltonian to describe the quantum phase transition between the quantum spin Hall insulator phase and a trivial band insulator phase depending on the relative strength of the spin-orbit coupling λ_{SO} and the sublattice potential Δ .

To better understand the Kane-Mele model, let us first calculate the corresponding Bloch Hamiltonian and then discuss its phase diagram. Performing a Fourier decomposition of the four-component spinor $\Psi = (a_{i\uparrow}, a_{i\downarrow}, b_{i\uparrow}, b_{i\downarrow})^T$ on the triangular Bravais lattice as before, we obtain the following Bloch Hamiltonian:

$$\mathcal{H}_{\text{KM}}(\mathbf{k}) = -(t/2)(\gamma(\mathbf{k}) \boldsymbol{\tau}_- \otimes \sigma_0 + \text{H.c.}) + \lambda_{\text{SO}} \tau_z \otimes (\mathbf{u}(\mathbf{k}) \cdot \boldsymbol{\sigma}) + \Delta \tau_z \otimes \sigma_0, \quad (5.25)$$

where, for simplicity, we consider the case $\lambda_{\text{R}}/t = 0$. As introduced before, the two vectors of Pauli matrices, $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)^T$ and $\boldsymbol{\tau} = (\tau_x, \tau_y, \tau_z)^T$ act on the spin and sublattice degrees of freedom, respectively, and $\sigma_0 = \tau_0 = \mathbb{1}$ denotes the 2×2 identity matrix. Furthermore, $\gamma(\mathbf{k})$ is a tight-binding function summed over the phase factors $\exp(i\mathbf{k} \cdot \mathbf{d}_a)$ corresponding to the three nearest neighbor atoms located at $\mathbf{d}_1, \mathbf{d}_2$, and \mathbf{d}_3 , and $\mathbf{u}(\mathbf{k})$ is a three-component real vector which consists of a sum over the phase factors $\exp[i\mathbf{k} \cdot (\mathbf{d}_j - \mathbf{d}_{j'})]$ connecting second-nearest neighbors:

$$\gamma(\mathbf{k}) \equiv \sum_{j=1}^3 e^{i\mathbf{k} \cdot \mathbf{d}_j}, \quad (5.26a)$$

$$\mathbf{u}(\mathbf{k}) \equiv \frac{i}{\mathcal{N}} \sum_{j,j'=1}^3 (\mathbf{d}_j \times \mathbf{d}_{j'}) e^{i\mathbf{k} \cdot (\mathbf{d}_j - \mathbf{d}_{j'})}. \quad (5.26b)$$

For the two-dimensional honeycomb lattice and using the normalization $\mathcal{N} = \sqrt{3}a^2/2$ we immediately see that $u_x(\mathbf{k}) = u_y(\mathbf{k}) = 0$, while $u_z(\mathbf{k})$ describes a spin-conserving spin-orbit interaction which respects the symmetries of graphene:

$$u_z(\mathbf{k}) = 4 \cos(3ak_x/2) \sin(\sqrt{3}ak_y/2) - 2 \sin(\sqrt{3}ak_y). \quad (5.27)$$

Note, however, that the gap generated by the spin-orbit interaction λ_{SO} is different from the gap that is generated by the staggered sublattice potential Δ : The spin-orbit interaction λ_{SO}

coupling to $\tau_z \otimes \sigma_z$ produces gaps with opposite sign at the two Dirac points \mathbf{K}_\pm , while the sublattice potential Δ couples to $\tau_z \otimes \sigma_0$ and produces gaps with the same sign at the Dirac points. Finally, for the sake of completeness, the full 4×4 Hamiltonian matrix takes the following form^{3,4}:

$$\mathcal{H}_{\text{KM}}(\mathbf{k}) = \begin{bmatrix} \lambda_{\text{SO}}u_z(\mathbf{k}) + \Delta & 0 & -t\gamma(\mathbf{k}) & 0 \\ 0 & -\lambda_{\text{SO}}u_z(\mathbf{k}) + \Delta & 0 & -t\gamma(\mathbf{k}) \\ -t\gamma(\mathbf{k})^* & 0 & -\lambda_{\text{SO}}u_z(\mathbf{k}) - \Delta & 0 \\ 0 & -t\gamma(\mathbf{k})^* & 0 & \lambda_{\text{SO}}u_z(\mathbf{k}) - \Delta \end{bmatrix}. \quad (5.28)$$

The electronic band structure of the Bloch Hamiltonian (5.25, 5.28) is obtained by a diagonalization of the Hermitian 4×4 matrix $\mathcal{H}(\mathbf{k})$. For the Kane-Mele Hamiltonian this leads to the following eigenvalues:

$$\mathcal{E}_{\pm\pm}(\mathbf{k}) = \pm \sqrt{t^2|\gamma(\mathbf{k})|^2 + (\lambda_{\text{SO}}|\mathbf{u}(\mathbf{k})| \pm \Delta)^2}. \quad (5.29)$$

Thus, in an inversion-symmetric system, *i.e.*, for $\Delta/t = 0$, the conduction and valence bands are two-fold degenerate as expected. In the absence of Rashba-type spin-orbit interactions, *i.e.*, for $\lambda_{\text{R}}/\lambda_{\text{SO}} = 0$, the bulk band structure is fully gapped with an energy gap of the size $\mathcal{E}_{\text{gap}} = 2|3\sqrt{3}\lambda_{\text{SO}} - \Delta|$ (Ref. 4). For $\Delta > 3\sqrt{3}\lambda_{\text{SO}}$ the bulk band gap is dominated by the staggered sublattice potential Δ , and the system becomes a trivial band insulator. However, in the opposite case $\Delta < 3\sqrt{3}\lambda_{\text{SO}}$, the Kane-Mele model describes the **quantum spin Hall insulator** phase which exhibits topologically protected edge states in finite-size ribbon geometries (see below). At the transition between the trivial insulator and the quantum spin Hall insulator, which occurs at $\Delta = 3\sqrt{3}\lambda_{\text{SO}}$, the bulk band gap closes, allowing the edge states to change their topological properties by switching time-reversed partner states [see also Fig. 5.8 (c, d)]. Finally, even in the presence of a finite Rashba spin-orbit interaction λ_{R} , the phase diagram of the Kane-Mele model contains a finite region which is adiabatically connected to the quantum spin Hall insulator phase at $\lambda_{\text{R}}/\lambda_{\text{SO}} = 0$ (cf. Fig. 5.6) as shown by Kane and Mele in Ref. 4.

5.2.5 Edge states of graphene nanoribbons

The different phases of the Kane-Mele model can also be understood from the behavior of the edge states in finite-size ribbons. As shown by Kane and Mele, in the quantum spin Hall insulator there exists a single pair of time-reversed eigenstates of the Kane-Mele Hamiltonian at each edge of the system³. Those gapless edge states are protected against small (non-magnetic) perturbations and against localization due to weak disorder, because time-reversal invariance prohibits mixing of Kramers' doublets and single-particle elastic backscattering is forbidden.

To develop a better understanding of those edge states, we therefore consider graphene nanoribbons with armchair and zigzag edges. Below we show numerically the appearance of edge states with a dispersion set by the spin-orbit interaction λ_{SO} or the next-nearest neighbor hopping amplitude t' . It turns out that those edge states become increasingly flat as we tune both λ_{SO} and t' towards the critical point $\lambda_{\text{SO}}/t = t'/t = 0$. A similar result was already obtained by Fujita *et al.* and Nakada *et al.* in 1996 who studied a semi-infinite graphene ribbon with zigzag edges, showing that it has a band of states which have zero energy and which are localized at the edges of the graphene ribbon^{160,161}.

In the first place, let us consider a graphene ribbon with zigzag edges as shown in Fig. 5.7, and let us assume that the ribbon has an infinite length in the longitudinal x direction, but a finite width of N unit cells in the transverse y direction. To illustrate the construction principle

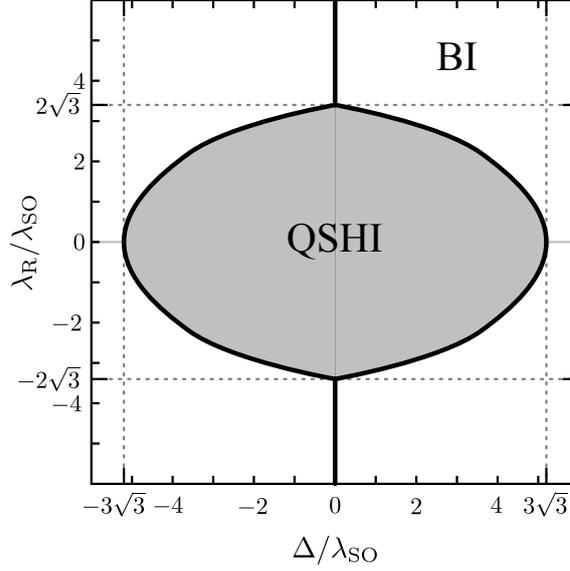


Figure 5.6. Phase diagram of the Kane-Mele model (5.23) as function of the staggered sublattice potential Δ and the Rashba-type spin-orbit interaction λ_R in units of λ_{SO} for $0 < \lambda_{SO}/t \ll 1$. For $\lambda_R/\lambda_{SO} = 0$, the quantum phase transition between the quantum spin Hall insulator (QSHI) and the trivial band insulator (BI) takes place at $|\Delta/\lambda_{SO}| = 3\sqrt{3} \approx 5.2$, while for $\Delta/\lambda_{SO} = 0$ the quantum phase transition takes place at $|\lambda_R/\lambda_{SO}| = 2\sqrt{3} \approx 3.5$. Note that in between there is a finite region in the phase diagram which is adiabatically connected to the quantum spin Hall insulator at $\lambda_R/\lambda_{SO} = 0$. Figure after Ref. 4.

for the tight-binding Hamiltonian in a ribbon geometry, we restrict the following derivation to the hopping term between nearest-neighbor atoms:

$$H_0 = -t \sum_{\langle i,j \rangle, \sigma} (c_{i\sigma}^\dagger c_{j\sigma} + \text{H.c.}). \quad (5.30)$$

Taking the two-atomic unit cell into account, we can rewrite that Hamiltonian in terms of the integer indices m and n which label the different unit cells of the semi-infinite graphene ribbon as defined in Fig. 5.7. As a result, we obtain:

$$H_0 = -t \sum_{m,n,\sigma} (a_{mn\sigma}^\dagger b_{mn\sigma} + a_{mn\sigma}^\dagger b_{m-1,n\sigma} + a_{mn\sigma}^\dagger b_{m,n-1,\sigma} + \text{H.c.}). \quad (5.31)$$

Since the graphene ribbon is assumed to be infinite in the longitudinal x direction, a Fourier decomposition of the operators in that direction yields

$$H_0 = -t \sum_{k,n,\sigma} [(1 + e^{ika}) a_{kn\sigma}^\dagger b_{kn\sigma} + a_{kn\sigma}^\dagger b_{k,n-1,\sigma} + \text{H.c.}]. \quad (5.32)$$

Note that a similar one-dimensional Fourier decomposition of operators can be performed also for the remaining three terms of the Kane-Mele Hamiltonian (5.23). The band structure of the graphene ribbon is then obtained from the resulting one-dimensional tight-binding Hamiltonian by means of exact diagonalization methods.

Fig. 5.8 shows the electronic band structure of a graphene ribbon with $N = 50$ unit cells in the transverse direction terminating in zigzag edges which was calculated in the above tight-binding approximation for different values of the spin-orbit coupling λ_{SO}/t and sublattice potential Δ/t . The bulk band gaps at the 1D projections $\bar{K}_\pm = \pi \pm \pi/3$ of the Dirac points \mathbf{K}_\pm

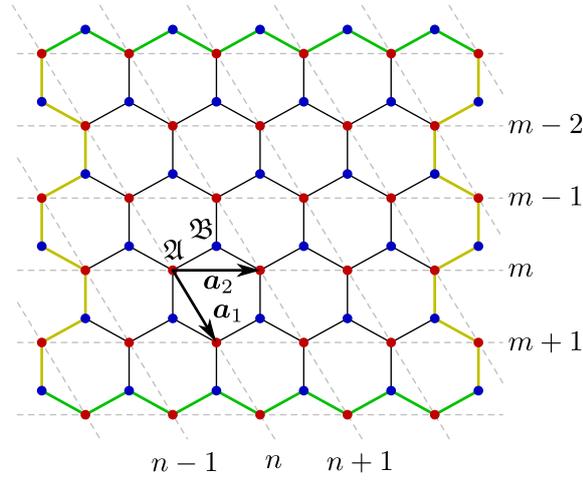


Figure 5.7. Sketch of a finite-size graphene ribbon with armchair (zigzag) edges which are indicated by yellow (green) lines in vertical (horizontal) direction. The two sublattices \mathfrak{A} and \mathfrak{B} of the honeycomb lattice are indicated by red and blue dots, respectively. The primitive lattice vectors \mathbf{a}_1 and \mathbf{a}_2 are chosen such that one can easily derive a tight-binding Hamiltonian for the zigzag edges along the horizontal direction. The indices m and n label the unit cells of the graphene lattice as the lattice vectors are given by $\mathbf{R} = m\mathbf{a}_1 + n\mathbf{a}_2$.

can be easily identified from their cone-like shapes. There are, however, two bands (indicated by red lines in Fig. 5.8) traversing the bulk band gap which connect the Dirac points and do not belong to the bulk states of the ribbon since they are localized at the edges of the ribbon^{160,161}. These edge states are helical states in the sense that electrons with opposite spin propagate in opposite directions. Interestingly, the edge states exist even in the limit $\lambda_{\text{SO}} \rightarrow 0$, where they become perfectly flat [cf. Fig. 5.8 (a)], leading to an enhanced local density of states at the Fermi energy associated with the zigzag edges. However, when electron-hole symmetry is broken by, e.g., a finite second-nearest hopping amplitude $t' \neq 0$, those states become dispersive with a Fermi velocity v_F^{edge} given by $t'a$ (Ref. 162). In that case, the band bending of the edge states opens up the possibility for elastic scattering into the bulk conduction and valence bands. However, the two-fold degeneracy of the electronic band structure is lifted upon breaking inversion symmetry by adding a staggered sublattice potential. For $\Delta/\lambda_{\text{SO}} < 3\sqrt{3}$ the system is adiabatically connected to the quantum spin Hall insulator at $\Delta/\lambda_{\text{SO}} = 0$ and shows similar edge states [cf. Fig. 5.8 (c)], while for $\Delta/\lambda_{\text{SO}} > 3\sqrt{3}$ the system is dominated by the band gap generated by the staggered sublattice potential and describes a trivial band insulator. In that case, the edge states do not cross the bulk band gap anymore, as shown in Fig. 5.8 (d).

A related effect occurs if chiral symmetry is broken by, e.g., finite spin-orbit interactions $\lambda_{\text{SO}}/t \neq 0$. In that case, the edge states become dispersive with a Fermi velocity of the edge states proportional to λ_{SO} , but the spin-orbit interactions also generate a bulk band gap $\mathcal{E}_{\text{gap}}^{\text{bulk}}$ whose size is determined by λ_{SO} as well [see Fig. 5.8 (b)]. Consequently, the edge states traversing the bulk band gap are protected against elastic scattering into the bulk bands. In general, if we consider a perturbation breaking chiral symmetry which also generates a bulk band gap $\mathcal{E}_{\text{gap}}^{\text{bulk}}$, but otherwise leaves the topological properties of the Kane-Mele model intact, then the edge states of the zigzag ribbon acquire a finite Fermi velocity which is of the order $v_F^{\text{edge}} \sim \mathcal{E}_{\text{gap}}^{\text{bulk}}/\Delta q$. Here, $\Delta q = 2\pi/3$ is the distance of the two Dirac points located at \mathbf{K}_{\pm} whose projection is given by $\bar{K}_{\pm} = \pi \pm \pi/3$ (see Fig. 5.8). In the case of spin-orbit interactions λ_{SO} , the bulk band gap is given by $\mathcal{E}_{\text{gap}}^{\text{bulk}} = 2 \cdot 3\sqrt{3} \lambda_{\text{SO}}$, and thus we can estimate $v_F^{\text{edge}} \sim 9\sqrt{3} \lambda_{\text{SO}}/\pi \approx$

$5\lambda_{\text{SO}}$ which is in agreement with the computed band structure for graphene ribbons with zigzag edges, as shown in Fig. 5.8.

Finally, note that similar edge states also occur in graphene ribbons with armchair edges, but in that case the 1D projections of the Dirac points $\bar{\mathbf{K}}_{\pm}$ are both at $\bar{K}_{\pm} = 0$. Hence, zero-energy states exist only at $k = 0$, even at the quantum critical point $\lambda_{\text{SO}}/t = 0$ (an analytical argument for the absence of flat edge states except for $k = 0$ is given below). Moreover, those edge states inherit their Fermi velocity from the bulk, *i.e.*, $v_F^{\text{edge}} = v_F^{\text{bulk}}$, and thus their properties cannot be easily manipulated in contrast to the zigzag case. In that sense, graphene ribbons with zigzag edges are much more interesting than armchairs edges, because they host perfectly flat edge states in the limit $\lambda_{\text{SO}}/t \rightarrow 0$.

5.2.6 Topological origin of zero-energy edge states

In this section, we briefly review the topological criteria for the existence of zero-energy edge states in graphene, as discussed by Ryu and Hatsugai¹⁸. The following discussion is mainly based on the topological criteria for the existence of topologically protected zero-energy bound states in the Su-Schrieffer-Heeger model which is discussed in appendix C. In particular, we show below that those zero-energy edge states form completely flat bands in graphene nanoribbons with zigzag edges. The topological arguments by Ryu and Hatsugai also provide a more general framework to discuss the zero-energy surface states in the 3D Fu-Kane-Mele model, which will be discussed in the next section.

In graphene, there are several ways to truncate the system, leading to different shapes of the edges for a graphene nanoribbon, such as zigzag, bearded, and armchair edges (see Fig. 5.9, left column). Starting from the bare Kane-Mele model for $\lambda_{\text{SO}}/t = \lambda_{\text{R}}/t = \Delta/t = 0$, including only hopping of electrons between nearest-neighbor sites, and neglecting the spin degree of freedom we can perform a Fourier decomposition of the spinors Ψ_i . As a result, we obtain a family of Hamiltonians which can be written in the form $H(k_x, k_y) = \mathbf{d}(k_x, k_y) \cdot \boldsymbol{\tau}$, where $\mathbf{d}(k_x, k_y) \in \mathbb{R}^3$. In the following, we choose coordinates such that the edge is always oriented along the y direction, so that the momentum k_y along the edge is a good quantum number, while we consider open boundary conditions in x direction (cf. Fig 5.9). Hence, we study a one-parameter family of one-dimensional tight-binding Hamiltonians which are parametrized by the momentum $k_y \in [-\pi/a, \pi/a]$. The vector \mathbf{d} for the different types of edges is given by¹⁸:

$$\mathbf{d}_{\text{zigzag}}(k_x, k_y) = (1 + \cos(k_y a) + \cos(k_x a - k_y a), \sin(k_x a - k_y a))^T, \quad (5.33a)$$

$$\mathbf{d}_{\text{bearded}}(k_x, k_y) = (1 + \cos(k_x a) + \cos(k_x a - k_y a), \sin(k_x a) + \sin(k_x a - k_y a))^T, \quad (5.33b)$$

$$\mathbf{d}_{\text{armchair}}(k_x, k_y) = (1 + \cos(k_x a) + \cos(k_x a + k_y a), \sin(k_x a) - \sin(k_x a + k_y a))^T, \quad (5.33c)$$

where, for simplicity, we have set $t = 1$. Importantly, note that $d_z(k_x, k_y) = 0$ for all momenta (k_x, k_y) within the 2D Brillouin zone due to a sublattice or chiral symmetry of the model, as discussed below. As a function of $k_x \in [-\pi/a, \pi/a]$ the vector $\mathbf{d}(k_x, k_y)$ forms a closed loop, because $k_x = \pm\pi/a$ are related by the reciprocal lattice vector $2\pi/a$ and thus $\mathbf{d}(k_x = \pi/a, k_y) = \mathbf{d}(k_x = -\pi/a, k_y)$. For a given value of $k_y \in [-\pi/a, \pi/a]$ one can calculate the corresponding winding number ν of $\hat{\mathbf{d}}$:

$$\nu \equiv \frac{1}{2\pi} \int_{-\pi/a}^{\pi/a} dk_x \epsilon_{\alpha\beta} \hat{d}_{\alpha}(k_x, k_y) \partial_{k_x} \hat{d}_{\beta}(k_x, k_y) \in \mathbb{Z}, \quad (5.34)$$

where $\epsilon_{\alpha\beta}$ denotes the totally anti-symmetric tensor with $\epsilon_{12} = 1$, and $\hat{\mathbf{d}}(k_x, k_y)$ is the normalized vector in the direction of $\mathbf{d}(k_x, k_y)$. Note that for a given momentum k_y this winding

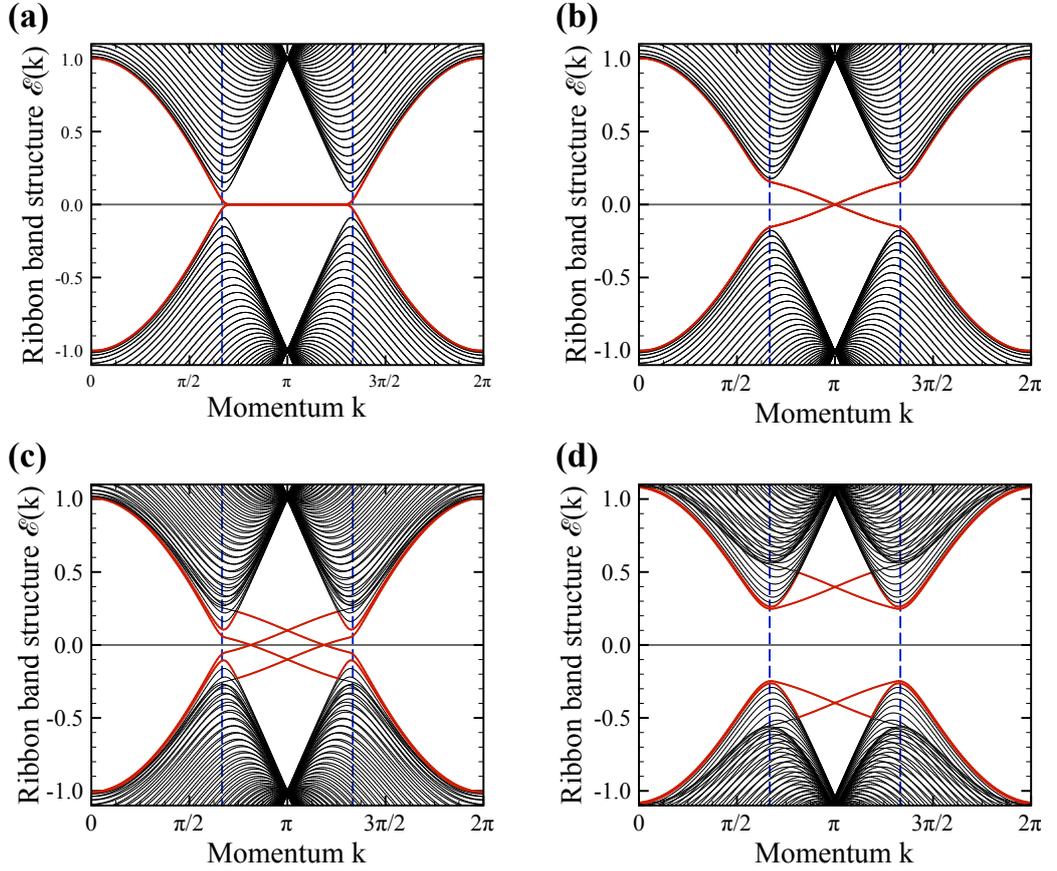


Figure 5.8. Electronic band structure of a graphene ribbon with $N = 50$ unit cells in the transverse direction terminating in zigzag edges. The blue dashed lines indicate the projections of the \mathbf{K}_{\pm} points onto $k = \pi \pm \pi/3$. The parameters of the Kane-Mele model have been chosen such that for $t = 1$ and $\lambda_R/t = 0$ the model (5.25) describes a quantum spin Hall insulator for $\Delta/\lambda_{SO} < 3\sqrt{3}$. **(a)** In the absence of spin-orbit coupling, *i.e.*, for $\lambda_{SO}/t = 0$, the zigzag ribbon exhibits perfectly flat zero-energy edge states. Although the bulk band gap $\mathcal{E}_{\text{gap}}^{\text{bulk}} = 2|3\sqrt{3}\lambda_{SO} - \Delta| = 0$, the band gap observed in the ribbon band structure stems from finite-size corrections, *i.e.*, $\mathcal{E}_{\text{gap}}^{\text{ribbon}} \sim \mathcal{O}(1/N)$. **(b)** In the presence of spin-orbit interactions (here, $\lambda_{SO}/t = 0.03$) those edge states become dispersive with a Fermi velocity given by $v_F \sim 9\sqrt{3}\lambda_{SO}/\pi \approx 5\lambda_{SO}$. **(c)** The two-fold degeneracy of the band structure is lifted upon breaking inversion symmetry by adding a staggered sublattice potential. For $\Delta/\lambda_{SO} < 3\sqrt{3}$ the system is adiabatically connected to the quantum spin Hall insulator at $\Delta/\lambda_{SO} = 0$ and shows similar edge states. **(d)** For $\Delta/\lambda_{SO} > 3\sqrt{3}$ the system is dominated by the band gap generated by the staggered sublattice potential and describes a trivial band insulator. Also, the edge states do not cross the bulk band gap anymore.

number ν counts how often the unit vector $\hat{\mathbf{d}}(k_x, k_y)$ winds around the unit circle as function of k_x . Furthermore, this winding number is intimately related to the first homotopy group of the unit circle, $\pi_1(S^1) = \mathbb{Z}$, and thus takes only integer values. The motivation and mathematical formulation of this winding number and its relation to the topologically protected edge states is explained in detail in appendix C for the Su-Schrieffer-Heeger.

The explicit calculation of the winding number for graphene ribbons with zigzag, bearded, and armchair edges then predicts zero-energy edge states in the case of zigzag edges for momenta $k_y \in [-\pi/a, -2\pi/3a] \cup [2\pi/3a, \pi/a]$ and for $k_y \in [-2\pi/3a, 2\pi/3a]$ in the case of bearded edges, while no zero-energy edge states are expected for the armchair edges except for $k_y = 0$. Explicit tight-binding calculations of the electronic band structure of graphene ribbons with different types of edges as shown in Fig. 5.9 have confirmed this scenario¹⁸. Moreover, since there is a finite region of momenta, where the zero-energy edge states occur, they are expected to form completely flat bands for zigzag and bearded edges, but not for armchair edges. In general, flat bands induce a sharp peak in the density of states at the Fermi energy which might trigger an instability in the presence of interactions. For example, electron-electron interactions can lead to a spontaneous magnetic polarization of the edges, as discussed below.

Now let us discuss how spin-orbit interactions, hopping between second-nearest neighbors, and the staggered sublattice potential affect the topological argument for the flatness of the edge states in graphene ribbons. As shown in appendix C, a necessary condition for a non-trivial winding of the $\mathbf{d}(\mathbf{k})$ vector is that the model Hamiltonian possesses a chiral symmetry so that $\mathbf{d}(\mathbf{k})$ lies within a 2D plane. Consequently, the existence of zero-energy edge states forming a perfectly flat band is due to the fact that the first homotopy group is non-trivial: $\pi_1(S^1) = \mathbb{Z}$. A close inspection of the Kane-Mele Hamiltonian (5.25) shows that chiral symmetry, *i.e.*, sublattice symmetry is represented by the operator $\hat{\Sigma} = \tau_z \otimes \sigma_0$, so that the condition $\{\hat{\Sigma}, \mathcal{H}_{\text{KM}}(\mathbf{k})\} \doteq 0$ immediately implies $\lambda_{\text{SO}}/t = \lambda_{\text{R}}/t = \Delta/t = 0$. From this we can infer that all three interactions break chiral symmetry, where the effective dispersion of the edge states is then set by the symmetry-breaking interaction strength. Mathematically, this is a consequence of the fact that the first homotopy group of higher-dimensional spheres is trivial:

$$\pi_1(S^n) = \begin{cases} \mathbb{Z} & n = 1 \\ 0 & n > 1 \end{cases} \quad (5.35)$$

In that sense, any perturbation breaking chiral symmetry also leads to a finite bandwidth of the edge states, and the question whether such a perturbation could lead to a spontaneously broken state with finite magnetization at the edges of a graphene ribbon is a matter of microscopic details. For example, for weak next-nearest neighbor hopping $t'/t \ll 1$, the bandwidth of the edge states is small compared to the bulk bandwidth, and we can hope to observe a spontaneously broken edge state, similar to case of perfectly flat edge states at $t'/t = 0$, but whether the kinetic energy or the interaction dominates depends on numbers.

5.2.7 Spontaneous ferromagnetism at the edges of graphene nanoribbons

In the last part of this section, we discuss the effects of electron-electron interactions on the boundary states of a single-layer graphene sheet. Since the localized edge states form a completely flat band, the local density of states develops a sharp peak and the electron compressibility diverges, which strongly affects the role of electron-electron interactions. For example, a number of *ab initio* calculations^{163–169} and mean-field analyses^{170,171} have predicted gapped edge states with ferromagnetic order, and also effective models^{172–174} show that close to half

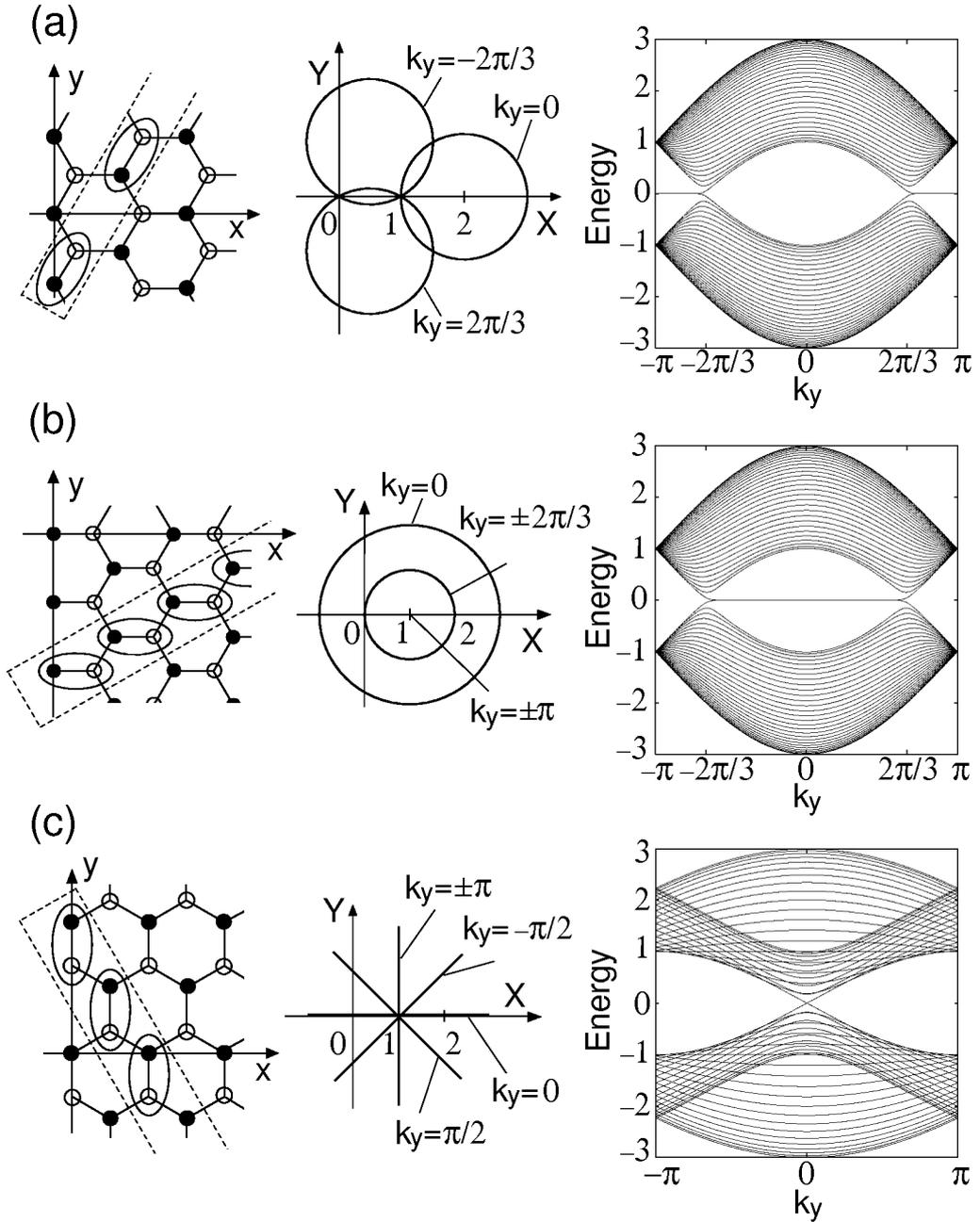


Figure 5.9. Loops of $d(k)$ in the two-dimensional $(X, Y) = (d_x, d_y)$ space and band structures of finite-size graphene ribbons with (a) zigzag, (b) bearded, and (c) armchair edges. In the left column, the ovals indicate the crystallographic unit cell with \mathfrak{A} and \mathfrak{B} sublattice, while the dotted squares indicate a choice for an enlarged unit cell for a finite-size ribbon. The loops in the middle panel corresponding to the one-parameter family of Bloch Hamiltonians $\mathcal{H}(k) = d(k) \cdot \tau$ are given by Eq. (5.33). Here, k_y is the conserved momentum along the graphene ribbon, and for simplicity we have taken $t = 1$. The existence of zero-energy edge states is predicted for the (a) zigzag edge and (b) bearded edge, but not for the (c) armchair edge due to the different winding numbers. Explicit tight-binding calculations for finite-size graphene ribbons (right column) confirm those zero-energy edge modes forming flat bands which are continuously connected to the bulk spectrum. Figure taken from Ref. 18.

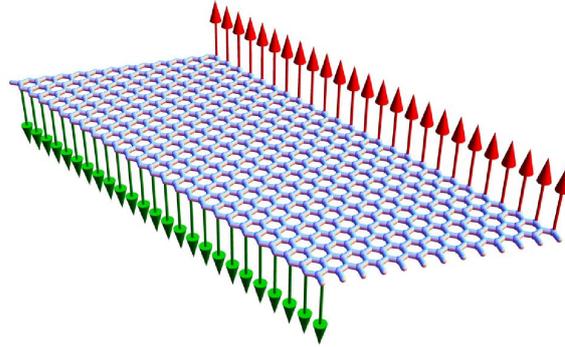


Figure 5.10. Schematic illustration of the ferromagnetic magnetization at the zigzag edges of a graphene ribbon. Figure taken from Ref. 122.

filling both short-ranged Hubbard interactions and long-ranged Coulomb interactions lead to full spin-polarized edge states.

A simple physical interpretation of this magnetic ground state is the following: electron-electron interactions try to minimize the interaction energy, thereby maximizing the distance between the electrons, which leads to a correlated ground state. First, in a graphene nanoribbon there is a mismatch between the number $N_{\mathfrak{A}, \mathfrak{B}}$ of \mathfrak{A} and \mathfrak{B} lattice sites, depending on the kind of edge termination. The most prominent example in the context of graphene are the aforementioned zigzag edges, where the outermost atom always corresponds to either the \mathfrak{A} sublattice or the \mathfrak{B} sublattice, but zero-energy edge states forming flat bands are present not only for zigzag edges, but for any boundary, except for pure armchair edges, where an equal number of \mathfrak{A} and \mathfrak{B} sites exist. Second, a local Hubbard interaction between spinful electrons in a flat band at half filling will favor a spin-polarized ground state with long-range ferromagnetic order (see Fig. 5.10), because local interactions of electrons with the same spin orientation are prohibited by the Pauli exclusion principle, thereby minimizing the total energy of the system. However, as the two zigzag edges of a graphene nanoribbon shown in Fig. 5.10 terminate on different sublattices, a local Hubbard interaction cannot cause any interaction between the states at the opposite edges. The only contribution is therefore a tunnel coupling between the edges which favors an anti-ferromagnetic alignment of the two edges¹⁷⁴. A similar argument can also be applied to long-range Coulomb interactions at half filling, where the system is charge neutral. Away from half-filling, the charged edges will interact with each other, but this inter-edge coupling is much weaker than the intra-edge coupling, so that it is expected not to alter qualitatively the ferromagnetic magnetization at the edges.

From a different point of view, one can understand the effect of electron-electron interactions on the edge states by considering the Kane-Mele model as an ideal quantum spin Hall insulator, where the left-moving states on the edge are correlated with spin-down ($\sigma = \downarrow$), while the right-movers have spin-up ($\sigma = \uparrow$). The correlation between the spin and the momentum gives rise to the notion of a **helical liquid**, in analogy to the helicity of massless Dirac fermions in a relativistic theory. When linearizing the band structure, we can describe the edge states in terms of a helical Luttinger liquid^{48,175}:

$$H_{\text{HLL}} = v_F \int dx (\psi_{R\uparrow}^\dagger(x) i \partial_x \psi_{R\uparrow}(x) - \psi_{L\downarrow}^\dagger(x) i \partial_x \psi_{L\downarrow}(x)), \quad (5.36)$$

where the right- and left-movers are denoted by $\psi_{R\uparrow}$ and $\psi_{L\downarrow}$, respectively. In that sense, the above Luttinger liquid theory can be considered “spinless” as spin and momentum are locked

and do not present independent quantum numbers. The usual single-particle backscattering term,

$$H_{\text{bs}} = g_{\text{bs}} \int dx (\psi_{R\uparrow}^\dagger(x)\psi_{L\downarrow}(x) + \text{H.c.}), \quad (5.37)$$

which opens up a mass gap in the spinless Luttinger liquid, is not allowed in the low-energy theory, as it is odd under time-reversal. There are, however, two time-reversal invariant interactions which are allowed, namely forward scattering:

$$H_{\text{fs}} = g_{\text{fs}} \int dx \psi_{R\uparrow}^\dagger(x)\psi_{R\uparrow}(x)\psi_{L\downarrow}^\dagger(x)\psi_{L\downarrow}(x), \quad (5.38)$$

and Umklapp scattering:

$$H_{\text{um}} = g_{\text{um}} \int dx (\psi_{R\uparrow}^\dagger(x)\psi_{R\uparrow}^\dagger(x+a)\psi_{L\downarrow}(x+a)\psi_{L\downarrow}(x) + \text{H.c.}), \quad (5.39)$$

where we have used a point splitting of the fermionic operators with the lattice constant a . The Umklapp process flips two spins simultaneously by scattering two left-movers with spin-down into two right-movers with spin-up and vice versa. On the other hand, the forward scattering term gives rise to a renormalized velocity u and a non-trivial Luttinger liquid parameter K , but the system remains gapless otherwise¹⁷⁵:

$$u = (v_F^2 - g_{\text{fs}}^2)^{1/2} \quad \text{and} \quad K = \left(\frac{v_F - g_{\text{fs}}}{v_F + g_{\text{fs}}} \right)^{1/2}. \quad (5.40)$$

Only the Umklapp scattering term converting two left-moving fermions into two right-moving ones can induce a mass gap close to or at half filling, *i.e.*, for $k_F = 0$. In the presence of Umklapp scattering, the standard Luttinger liquid Hamiltonian after bosonization takes the following form¹⁷⁵:

$$H = \frac{1}{2\pi} \int dx \left[\frac{u}{K} (\partial_x \phi(x))^2 + uK (\partial_x \theta(x))^2 \right] + \frac{g_{\text{um}}}{2(\pi a)^2} \int dx \cos(\sqrt{8}\phi), \quad (5.41)$$

where $\phi(x) \equiv \phi_R(x) + \phi_L(x)$ and $\theta(x) \equiv \phi_R(x) - \phi_L(x)$ are two bosonic fields describing the charge and current excitations of the system. This Hamiltonian is also known as **sine-Gordon Hamiltonian** in the field theory literature. Physically, the effect of the cosine is the following: In contrast to the quadratic terms, which let the field $\phi(x)$ fluctuate, the cosine term would like to lock the field $\phi(x)$ into one of the minima of the cosine, so that there will be a competition between the quadratic part and the cosine. The renormalization group analysis shows that the Umklapp term becomes relevant for $K < K_c = 1/2$, where the field ϕ is locked into one of the minima of the cosine¹⁷⁵. Hence, the $\phi(x)$ field orders, and we get into a massive phase with a gap of $\Delta \approx a^{-1} g_{\text{um}}^{1/(2-4K)}$, where the system spontaneously breaks time-reversal symmetry by developing long-range ferromagnetic order at half filling and zero temperature. For the helical Luttinger liquid we observe that the value of the Luttinger liquid parameter K can be tuned by changing the Fermi velocity v_F [cf. Eq. (5.40)]. In particular, by making v_F sufficiently small we obtain $v_F \approx g_{\text{fs}}$ and thus $K < K_c = 1/2$ which marks the transition to the ferromagnetic edge channels.

5.2.8 Discussion

To conclude, we have studied the Kane-Mele Hamiltonian as a model for 2D topological insulators with topologically protected edge states, and we have discussed the topological criterion

of chiral symmetry to predict zero-energy edge states. Furthermore, we have shown that those edge states form completely flat bands in the limit $\lambda_{\text{SO}} \rightarrow 0$, leading to a sharp peak in the local density of states. Consequently, interactions become more relevant, and the strongly interacting edge states are prone to many-body instabilities leading to a spontaneous ferromagnetic order at the boundaries of graphene. Hence, generalizing the Kane-Mele model on the 2D graphene lattice to the 3D diamond lattice we expect a similar mechanism on the 2D surfaces of such a 3D topological insulator as well, namely in the context of chiral symmetry breaking which describes the spontaneous gap generation of 2D Dirac fermions in the presence of long-range Coulomb interactions. In the remainder of this chapter, we investigate possible routes towards such a surface state of a 3D topological insulator with a spontaneously generated mass due to interactions.

5.3 The Fu-Kane-Mele model on the diamond lattice

In the previous section, we have studied the Kane-Mele Hamiltonian (5.23) on the graphene lattice which was introduced as a model Hamiltonian in the context of 2D topological insulators in 2005 by Kane and Mele^{3,4}. Soon after, this model was generalized by Fu, Kane, and Mele to three spatial dimensions, introducing the so-called **Fu-Kane-Mele Hamiltonian** on the diamond lattice as a 3D analogue of the Kane-Mele model^{7,8}:

$$\begin{aligned}
 H_{\text{FKM}} = & -t \sum_{\langle i,j \rangle, \sigma} (c_{i\sigma}^\dagger c_{j\sigma} + \text{H.c.}) + i\lambda_{\text{SO}} \sum_{\langle i,j \rangle, \sigma, \sigma'} c_{i\sigma}^\dagger (\boldsymbol{\nu}_{ij} \cdot \boldsymbol{\sigma})_{\sigma\sigma'} c_{j\sigma'} \\
 & + \Delta \sum_{i, \sigma} \xi_i c_{i\sigma}^\dagger c_{i\sigma} - t' \sum_{\langle\langle i,j \rangle\rangle, \sigma} (c_{i\sigma}^\dagger c_{j\sigma} + \text{H.c.}).
 \end{aligned} \tag{5.42}$$

Note that the Kane-Mele Hamiltonian and the Fu-Kane-Mele Hamiltonian are quite similar, but differ in dimensionality due to the underlying lattices and crystal symmetries. For example, on the graphene lattice the vector $\boldsymbol{\nu}_{ij}$ has the form $(0, 0, \nu_{ij}^z)^T$ due to the symmetries of the lattice. Thus, the spin-orbit interaction of the electrons conserves the spin. In contrast, on the diamond lattice all elements of the vector $\boldsymbol{\nu}_{ij}$ have non-zero components $\nu_{ij}^\alpha = \pm 1$ (with $\alpha \in \{x, y, z\}$), therefore leading to a mixing of spin-up and spin-down states. Furthermore, the Rashba-type spin-orbit interaction λ_{R} , which in graphene is generated due to the lack of mirror symmetry in the presence of, *e.g.*, a substrate, is absent on the diamond lattice due to full inversion symmetry. As a consequence, the two models also have different bulk phase diagrams: Depending on the model parameters, the Kane-Mele Hamiltonian exhibits both a trivial band insulator phase and quantum spin Hall insulator phase, whereas the Fu-Kane-Mele Hamiltonian exhibits only a semimetallic phase. To make this model either a strong or a weak topological insulator, we have to include additional anisotropies in the overlap parameters, as discussed below. Moreover, we also include a spin-independent hopping term t' between next-nearest neighboring atoms as leading-order correction to the usual hopping term. This term gives rise to an energy correction which is identical for both sublattices of the diamond lattice and destroys the particle-hole symmetry of the conduction and valence bands. Similar to the edge states in graphene ribbons, where the bands formed by the edge states also become dispersive with a Fermi velocity $v_F = at'$ when particle-hole symmetry is broken by second-nearest neighbor hopping^{155,162}, we expect similar effects of this term on the surface states in a diamond slab geometry.

In the following, we first consider the crystal symmetries of the diamond lattice, and then discuss the electronic band structure of the Fu-Kane-Mele Hamiltonian. In slab geometries,

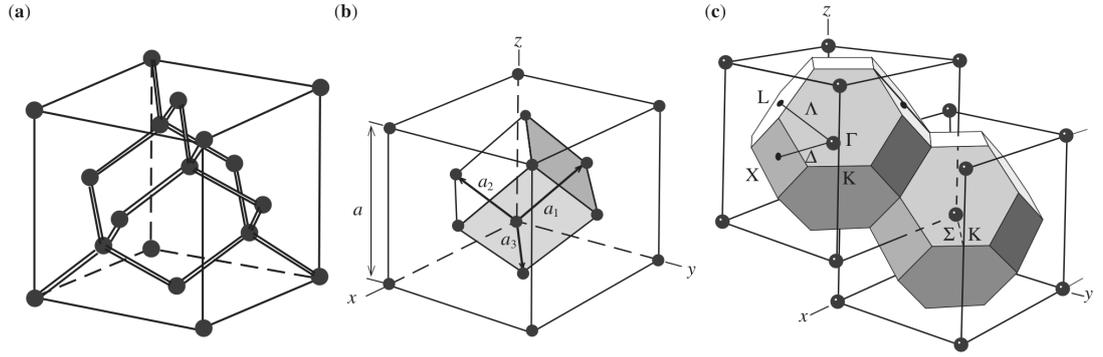


Figure 5.11. (a) The crystal structure of the diamond lattice. (b) The face-centered cubic lattice with a set of primitive lattice vectors. (c) The reciprocal lattice of the fcc lattice and the first Brillouin zone. Points of high symmetry are denoted by $\Gamma = (0, 0, 0)^T$, $X = \frac{2\pi}{a}(1, 0, 0)^T$, $K = \frac{2\pi}{a}(\frac{3}{4}, \frac{3}{4}, 0)^T$, and $L = \frac{2\pi}{a}(\frac{1}{2}, \frac{1}{2}, \frac{1}{2})^T$, while the three high-symmetry directions $[100]$, $[110]$, and $[111]$ in the Brillouin zone are denoted by Δ , Σ and Λ .

we find surface states which become perfectly flat at the bulk quantum critical point $\lambda_{\text{SO}}/t = t'/t = 0$. Those flat surface bands can be explained by a straightforward generalization of the **topological criteria** given for the zero-energy edge states of the Kane-Mele model on the graphene lattice, as discussed in the previous section. We also show that they are related to **nodal lines** in the bulk of the system, where the bulk band gap vanishes. Moreover, the surface bands in principle open up the possibility to find a surface state with a spontaneously generated mass term, thereby creating a gapped system as discussed in section 5.1. Concerning the relevant interaction parameter $\alpha = e^2/(\hbar v_F^{\text{surf}})$ [cf. Eq. (5.7)] our strategy is to decrease the Fermi velocity v_F^{surf} of the surface Dirac fermions which can be achieved by approaching the bulk quantum critical point. The appropriate knob here to control the Fermi velocity is the spin-orbit coupling parameter λ_{SO} which has been demonstrated experimentally via chemical substitution in bismuth-based compounds⁷², but can in principle also be achieved in mercury telluride systems by replacing mercury atoms with telluride atoms⁶. However, one also has to investigate to what extent the dielectric constant ϵ characterizing the bulk of the sample is affected by the adiabatic change in the spin-orbit coupling. In particular, we show by explicitly calculating the polarization function that the dielectric constant diverges as λ_{SO} vanishes. As a consequence of this divergence, the spontaneous mass generation on the surface of a strong topological insulator is hampered by screening due to the presence of the aforementioned bulk nodal lines.

5.3.1 The diamond lattice

Since the diamond lattice shown in Fig. 5.11 is very similar to the zinc-blende lattice discussed in section 3.3.1, we only briefly mention the basic results. First, the **diamond lattice** is a face-centered cubic (fcc) Bravais lattice with a two-atomic unit cell, and those identical atoms are located at $\mathbf{r}_1 = (0, 0, 0)^T$ and $\mathbf{r}_2 = \frac{a}{4}(1, 1, 1)^T$, giving rise to two sublattices denoted \mathfrak{A} and \mathfrak{B} . The primitive lattice vectors \mathbf{a}_1 , \mathbf{a}_2 , and \mathbf{a}_3 and the corresponding primitive reciprocal lattice vectors \mathbf{b}_1 , \mathbf{b}_2 , and \mathbf{b}_3 defined by the relation $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$ are given by Eqs. (3.16) and (3.18). The Brillouin zone of the diamond lattice forms a truncated octahedron, as shown in 5.11 (c). Since the reciprocal lattice vectors are obtained from the direct lattice vectors, the symmetries of the Brillouin zone are also determined by the symmetries of the crystal lattice.

The two atoms in the unit cell of the diamond lattice are equivalent, and this has impor-

class	symmetries of the space group #227 / O_h^7 / $Fd\bar{3}m$
$\{E\}$	identity
$\{3C_2\}$	two-fold rotations of 180° about each of the $[100]$, $[010]$, and $[001]$ axes
$\{6S_4\}$	two four-fold improper clockwise and counterclockwise rotations of 90° about the $[100]$, $[010]$ and $[001]$ axes
$\{6\sigma_d\}$	reflections on each of the (110) , $(\bar{1}\bar{1}0)$, (101) , $(10\bar{1})$, (011) , and $(01\bar{1})$ planes
$\{8C_3\}$	three-fold clockwise and counterclockwise rotations of 120° about each of the $[111]$, $[\bar{1}\bar{1}1]$, $[1\bar{1}\bar{1}]$, and $[11\bar{1}]$ axes
$\{i\}$	inversion
$\{3\sigma_h\}$	reflections on each of the (100) , (010) , and (001) planes
$\{6C_4\}$	two four-fold clockwise and counterclockwise rotations of 90° about the $[100]$, $[010]$, and $[001]$ axes
$\{6C_2'\}$	two-fold rotations of 180° about each of the $[110]$, $[101]$, and $[011]$ axes
$\{8S_6\}$	two three-fold improper clockwise and counterclockwise rotations of 60° about $[111]$, $[\bar{1}\bar{1}1]$, $[1\bar{1}\bar{1}]$, and $[11\bar{1}]$ axes

Table 5.2. The symmetry group O_h^7 of the diamond lattice has the same symmetry operators as the point group O_h of the cube. The first five symmetry classes are the same as those of T_d , while the other five are obtained from the first five by multiplication with the inversion operation.

tant consequences compared to the zinc-blende lattice (cf. chapter 3): The **space group #227** of the diamond lattice, also denoted by O_h^7 or $Fd\bar{3}m$ in international notation, is different from the space group T_d^2 ($F\bar{4}3m$) of the zinc-blende lattice discussed in section 3.3.1. Here, $Fd\bar{3}m$ describes a face-centered cubic lattice with diamond glide planes normal to the $\langle 100 \rangle$ axes, three-fold improper rotations about the $\langle 111 \rangle$ axes, and reflections on the $\{110\}$ mirror planes. Moreover, the space group O_h^7 is **non-symmorphic**, because it contains three glide planes defined by $x = \frac{a}{8}$, $y = \frac{a}{8}$, and $z = \frac{a}{8}$. For example, the plane defined by $x = \frac{a}{8}$ is a glide plane, because the diamond lattice is invariant under a translation by $\frac{a}{4}(0, 1, 1)^T$ followed by a reflection on this plane. Choosing the origin at the center of the line joining the two identical atoms we observe that the crystal structure is invariant under spatial inversion with respect to this origin, showing that the point group contains the **inversion** as point-group operation as well.

Therefore, the factor group of the diamond lattice is isomorphic to the point group generated from the point group T_d of the zinc-blende lattice by adding the inversion operation. Consequently, this **point group** consists of 48 elements and is denoted by O_h . The 48 elements of this point group are divided into 10 classes $\{E\}$, $\{3C_2\}$, $\{6S_4\}$, $\{6\sigma_d\}$, $\{8C_3\}$, $\{i\}$, $\{3\sigma_h\}$, $\{6C_4\}$, $\{6C_2'\}$, and $\{8S_6\}$ which are listed in Table 5.2 for reference. Hence, O_h has ten **irreducible representations** which in atomic physics are denoted by A_{1g} , A_{2g} , E_g , T_{1g} , T_{2g} , A_{1u} , A_{2u} , E_u , T_{1u} , and T_{2u} . Here, “u” and “g” refer to the even (odd) parity of the irreducible representations under those symmetry operations, respectively. As mentioned before, the irreducible representations are often classified by their symmetries at the center of the Brillouin zone, i.e., the Γ point, and are thus labeled by Γ_1^\pm , Γ_2^\pm , Γ_3^\pm , Γ_4^\pm , and Γ_5^\pm , where the superscript “ \pm ” denotes the even/odd parity of the representation⁷⁴.

5.3.2 Electronic band structure of the Fu-Kane-Mele model

The electronic band structure of the Fu-Kane-Mele model may be calculated in a similar way as for the Kane-Mele model following the general tight-binding approach. First, we decompose

the position of an atom in the diamond lattice into $r_{jl} = \mathbf{R}_j + \mathbf{r}_l$, where \mathbf{R}_j denotes the position of the unit cell within the fcc lattice, and \mathbf{r}_l is the position of the atom within the unit cell. For the diamond lattice we have $l = 1, 2$ only. In analogy to the graphene lattice, we consider two sublattices \mathfrak{A} ($l = 1$) and \mathfrak{B} ($l = 2$) for the diamond lattice with annihilation and creation operators a, a^\dagger and b, b^\dagger acting on the \mathfrak{A} and \mathfrak{B} sublattice, respectively. As before, we define the four-component spinor $\Psi_i = (a_{i\uparrow}, a_{i\downarrow}, b_{i\uparrow}, b_{i\downarrow})^T$ comprising both the sublattice and spin degrees of freedom of the electrons, and we define the Fourier transform of Ψ_i as in Eq. (5.16). The straightforward Fourier transformation of the Fu-Kane-Mele Hamiltonian on the diamond lattice (5.42) then results in the following Bloch Hamiltonian:

$$\begin{aligned} \mathcal{H}_{\text{FKM}}(\mathbf{k}) \equiv & -(t/2)(\gamma(\mathbf{k}) \tau_+ \otimes \sigma_0 + \text{H.c.}) + \lambda_{\text{SO}} \tau_z \otimes (\mathbf{u}(\mathbf{k}) \cdot \boldsymbol{\sigma}) \\ & + \Delta \tau_z \otimes \sigma_0 - t' \gamma'(\mathbf{k}) \tau_0 \otimes \sigma_0. \end{aligned} \quad (5.43)$$

As introduced before, the two vectors of Pauli matrices, $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)^T$ and $\boldsymbol{\tau} = (\tau_x, \tau_y, \tau_z)^T$ act on the spin and sublattices degrees of freedom, respectively, $\sigma_0 = \tau_0 = \mathbb{1}$ is the 2×2 identity matrix, and $\tau_{\pm} = \tau_x \pm i\tau_y$ denote raising and lowering operators describing hopping between the \mathfrak{A} and \mathfrak{B} sublattice. As for the Kane-Mele model, we can calculate the matrix structure of the Bloch Hamiltonian which takes the following form:

$$\mathcal{H}_{\text{FKM}}(\mathbf{k}) = \begin{bmatrix} \lambda_{\text{SO}} u_z(\mathbf{k}) + \Delta & \lambda_{\text{SO}} u_-(\mathbf{k}) & -t\gamma(\mathbf{k}) & 0 \\ \lambda_{\text{SO}} u_+(\mathbf{k}) & -\lambda_{\text{SO}} u_z(\mathbf{k}) + \Delta & 0 & -t\gamma(\mathbf{k}) \\ -t\gamma(\mathbf{k})^* & 0 & -\lambda_{\text{SO}} u_z(\mathbf{k}) - \Delta & -\lambda_{\text{SO}} u_-(\mathbf{k}) \\ 0 & -t\gamma(\mathbf{k})^* & -\lambda_{\text{SO}} u_+(\mathbf{k}) & \lambda_{\text{SO}} u_z(\mathbf{k}) - \Delta \end{bmatrix}, \quad (5.44)$$

where we have omitted the diagonal elements $-t' \gamma'(\mathbf{k}) \tau_0 \otimes \sigma_0$ for clarity. Here, $\gamma(\mathbf{k})$, $\gamma'(\mathbf{k})$, and $\mathbf{u}(\mathbf{k})$ are tight-binding functions which are summed over nearest and second-nearest neighbors on the diamond lattice, respectively, and $u_{\pm}(\mathbf{k}) \equiv u_x(\mathbf{k}) \pm iu_y(\mathbf{k})$. First, the hopping between nearest-neighboring and next-nearest neighboring atoms is described in terms of $\gamma(\mathbf{k})$ and $\gamma'(\mathbf{k})$, respectively:

$$\gamma(\mathbf{k}) \equiv \sum_{j=1}^4 e^{i\mathbf{k} \cdot \mathbf{d}_j}, \quad \gamma'(\mathbf{k}) \equiv \sum_{j \neq j'=1}^4 e^{i\mathbf{k} \cdot (\mathbf{d}_j - \mathbf{d}_{j'})}. \quad (5.45)$$

On the diamond lattice, $\gamma(\mathbf{k})$ and $\gamma'(\mathbf{k})$ are given by:

$$\gamma(\mathbf{k}) = 4 \left[\cos\left(\frac{ak_x}{4}\right) \cos\left(\frac{ak_y}{4}\right) \cos\left(\frac{ak_z}{4}\right) - i \sin\left(\frac{ak_x}{4}\right) \sin\left(\frac{ak_y}{4}\right) \sin\left(\frac{ak_z}{4}\right) \right], \quad (5.46a)$$

$$\gamma'(\mathbf{k}) = 4 \left[\cos\left(\frac{ak_x}{2}\right) \cos\left(\frac{ak_y}{2}\right) + \cos\left(\frac{ak_y}{2}\right) \cos\left(\frac{ak_z}{2}\right) + \cos\left(\frac{ak_z}{2}\right) \cos\left(\frac{ak_x}{2}\right) \right]. \quad (5.46b)$$

Second, $\mathbf{u}(\mathbf{k})$ is a three-component real vector which describes the effective spin-orbit interaction generated by a spin-dependent hopping term connecting second-nearest neighbors. On the diamond lattice all components of $\mathbf{u}(\mathbf{k})$ are non-vanishing:

$$u_x(\mathbf{k}) \equiv i \sum_{i,j=1}^4 (\boldsymbol{\nu}_{ij})_x e^{i\mathbf{k} \cdot (\mathbf{d}_i - \mathbf{d}_j)} = 4 \sin\left(\frac{ak_x}{2}\right) \left[\cos\left(\frac{ak_y}{2}\right) - \cos\left(\frac{ak_z}{2}\right) \right], \quad (5.47)$$

and the remaining components $u_y(\mathbf{k})$ and $u_z(\mathbf{k})$ are obtained from $u_x(\mathbf{k})$ by cyclic permutation of the indices x, y , and z . Note that ν_{ij} is defined in such a way that its components are of unit magnitude:

$$\nu_{ij} \equiv \frac{1}{\mathcal{N}} \left[\mathbf{d}_{ij}^{(1)} \times \mathbf{d}_{ij}^{(2)} \right] = -\frac{1}{\mathcal{N}} (\mathbf{d}_i \times \mathbf{d}_j) \quad \text{with} \quad \mathcal{N} \equiv \frac{a^2}{8}, \quad (5.48)$$

where $\mathbf{d}_{ij}^{(1,2)}$ denote the two bonds \mathbf{d}_i and \mathbf{d}_j traversed by electron when moving from site i to j . As mentioned before, the Kane-Mele Hamiltonian on the graphene lattice and the Fu-Kane-Mele Hamiltonian on the diamond lattice are essentially equivalent, thus leading to the same Bloch Hamiltonian in the tight-binding approach. However, in the 2D case of graphene, the spin-orbit interaction is spin-conserving, *i.e.*, $u_x(\mathbf{k}) = u_y(\mathbf{k}) = 0$, whereas in the 3D case of the diamond lattice the different spin-up and spin-down states mix.

The four energy bands of the Fu-Kane-Mele Hamiltonian can be calculated by diagonalizing the corresponding Bloch Hamiltonian (5.43). As a result, we obtain:

$$\mathcal{E}_{\pm\pm}(\mathbf{k}) = -t'\gamma'(\mathbf{k}) \pm \sqrt{t^2|\gamma(\mathbf{k})|^2 + (\lambda_{\text{SO}}|\mathbf{u}(\mathbf{k})| \pm \Delta)^2}. \quad (5.49)$$

Apparently, the presence of a hopping term between next-nearest neighbor atoms shifts the Fermi level of the model and breaks particle-hole symmetry, while the gaps induced by the spin-orbit interaction λ_{SO} and the staggered sublattice potential Δ compete. Note that in an inversion-symmetric system, *i.e.*, for $\Delta/t = 0$, each band is doubly degenerate, and the conduction and valence bands meet at the three equivalent X points in the Brillouin zone, $X_\alpha = (2\pi/a)\mathbf{e}_\alpha$ for $\alpha \in \{x, y, z\}$. To lift this degeneracy and to make the Fu-Kane-Mele model a topological insulator, we have to include anisotropic modulations of the hopping amplitude t on the four nearest-neighbor bonds, *i.e.*, we set $t \rightarrow t_j = t + \delta t_j$ ($j = 1, \dots, 4$) which lowers the crystal symmetries of the diamond structure⁷. As a consequence, the tight-binding function $\gamma(\mathbf{k})$ describing the hopping between nearest neighbor atoms is modified as well,

$$\gamma(\mathbf{k}) \equiv \sum_{j=1}^4 \left(1 + \frac{\delta t_j}{t} \right) e^{i\mathbf{k} \cdot \mathbf{d}_j}, \quad (5.50)$$

while we neglect the changes of the next-nearest-neighbor tight-binding function $\gamma'(\mathbf{k})$, because usually the overlap parameter t' is much smaller than t , *i.e.*, $t'/t \ll 1$. Also note that the (gapped) bulk Dirac points defined by the two conditions $|\gamma(\mathbf{k})| \doteq 0$ and $|\mathbf{u}(\mathbf{k})| \doteq 0$ are shifted within the Brillouin zone as function of the hopping amplitude modulations δt_j .

Fig. 5.12 illustrates the effect of the various model parameters on the electronic band structure $\mathcal{E}_{\pm\pm}(\mathbf{k})$ of the Fu-Kane-Mele model. Throughout this section we consider all model parameters to be equal to zero unless stated otherwise, except for the hopping amplitude which is usually set to unity. Panel (a) shows the band structure of the pristine Fu-Kane-Mele model in the absence of spin-orbit interactions λ_{SO} , modulated hopping amplitude δt_j , and a staggered sublattice potential Δ . Apparently, there exists a critical line of high-symmetry between the X and W points in the Brillouin zone, where the conduction and valence band have the same energy, leading to a semimetal. By including spin-orbit interactions, the degeneracy of conduction and valence bands is lifted, but the system remains a semimetal, because the bands still touch at the high-symmetry point X [cf. panel (b)]. If we further consider anisotropic hopping amplitudes, as shown in panel (c) for $\delta t_1/t = 0.4$, this bulk Dirac point is shifted away from the X point. Note, however, that the conduction and valence bands still touch at some other point \mathbf{k}^* within the bulk Brillouin zone which is determined by the relation $|\gamma(\mathbf{k}^*)| \doteq 0$ (not shown in Fig. 5.12). The gap induced by the staggered sublattice potential Δ creates a finite

bulk band gap and makes the system an ordinary band insulator [cf. panel (d)]. In the presence of both sublattice potential Δ , which breaks the inversion symmetry of the model, and spin-orbit interactions, both gaps compete with each other and lift the two-fold degeneracy of the conduction and valence bands [cf. Eq. (5.49)]. In particular, there exists a critical line in the parameter space defined by $\Delta/\lambda_{\text{SO}} = 8$ which separates a topologically non-trivial phase from the trivial band insulator, similar to the phase diagram obtained for the Kane-Mele model on the graphene lattice (cf. Fig. 5.6). To be specific, for a given sublattice potential Δ we observe a topologically non-trivial phase for $\lambda_{\text{SO}} > \lambda_{\text{SO}}^c = \Delta/8$ [cf. panel (e)], but a trivial band insulator for $\lambda_{\text{SO}} < \lambda_{\text{SO}}^c = \Delta/8$ [cf. panel (f)].

From the bulk density of states $\nu(\omega)$ plotted in Fig. 5.13 as function of the energy ω/t we also see that the line of bulk Dirac points, which only exists in the absence of spin-orbit interactions, leads to a linear dependence $\nu(\omega) \sim |\omega|$ close to the Fermi level [cf. panels (a, c, d)]. On the other hand, we observe that the degeneracy of the valence and conduction bands along this nodal line is lifted by including spin-orbit interactions. However, the bands still touch at the isolated three-dimensional Dirac points X . Consequently, the density of states vanishes quadratically, *i.e.*, $\nu(\omega) \sim \omega^2$ [cf. panel (b)]. Furthermore, the second-nearest neighbor hopping term t' breaks particle-hole symmetry as can be seen from the asymmetric density of states in panels (d, f). Finally, if we consider the situation, where hopping along one of the bonds is stronger than along the others ($\delta t_1/t = 0.4$), and take spin-orbit interaction into account ($\lambda_{\text{SO}}/t = 0.125$) we observe that a global band gap is opened in panels (e, f), in agreement with the electronic bulk band structure shown in Fig. 5.12 (e). In principle, this bulk band gap opens up the possibility for the existence of topologically protected surface states, and those surface states are of interest in the scenario of chiral symmetry breaking on the surfaces of a strong topological insulator, as discussed before.

To explain the phase diagram of the Fu-Kane-Mele model shown in Fig. 5.14 let us briefly summarize the results of Refs. 7,8. Most importantly, the different phases of the Fu-Kane-Mele model can be distinguished by the four \mathbb{Z}_2 invariants $(\nu_0; \nu_1\nu_2\nu_3)$, which have been introduced in section 2.5 and which have been discussed for the minimal HgTe model in more detail in section 3.6.6. In the following, we review both the weak and strong topological insulator phases with $\nu_0 = 0$ and $\nu_0 = 1$, respectively.

For $\nu_0 = 0$, the Fu-Kane-Mele model describes a **weak topological insulator**, and this quantum state of matter is characterized by an even number of Dirac fermions on each surface. It can be interpreted as a 3D sample consisting of layers of 2D quantum spin Hall states which are stacked along the $[\nu_1\nu_2\nu_3]$ direction, where ν_1 , ν_2 , and ν_3 take the role of Miller indices for the planes showing the quantum Hall phases. As pointed out by Fu, Kane, and Mele, the presence or absence of topologically protected surface states is rather involved in the case of $\nu_0 = 0$. It turns out that the number of Dirac fermions on a particular surface of the crystal depends very much on the translational symmetries of the lattice in the sense that upon doubling the unit cell two Dirac points will be folded back onto one another. A weak periodic potential or disorder would then induce a band gap, thus localizing and eliminating the topological surface states.

The quantum states with $\nu_0 = 1$, on the other hand, are much more robust against disorder and are therefore called **strong topological insulators**. In contrast to weak topological insulators, a strong topological insulator hosts an odd number of Dirac fermions on each of its surfaces, with partner Dirac fermions residing on opposite surfaces of the sample. In the simplest case, the surface Fermi surface encloses a single Dirac point. Due to the π Berry phase of a single Dirac fermion, the surface state is immune to non-magnetic disorder and cannot be localized by weak disorder. In other words, a strong 3D topological insulator exhibits a 2D

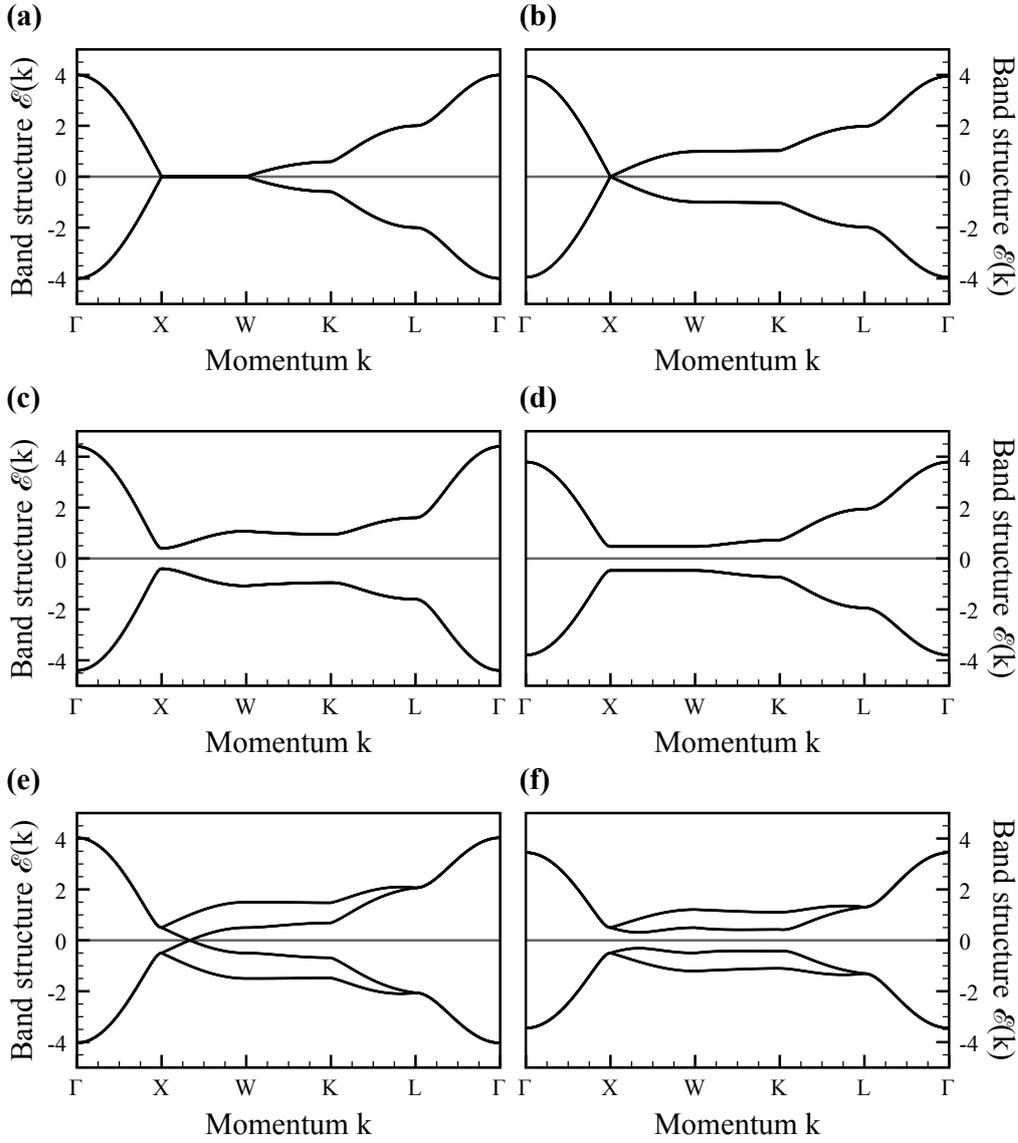


Figure 5.12. Particle-hole symmetric energy bands $\mathcal{E}_{\pm\pm}(\mathbf{k})$ [cf. Eq. (5.49)] of the Fu-Kane-Mele Hamiltonian (5.43) as function of the crystal momentum \mathbf{k} with $t = 1$. **(a)** For $\lambda_{\text{SO}}/t = \delta t_j/t = \Delta/t = 0$ both conduction and valence bands are doubly degenerate, and there is a nodal line with zero energy connecting X and W . **(b)** The degeneracy of valence and conduction bands along that nodal line is lifted by including a finite spin-orbit coupling parameter ($\lambda_{\text{SO}}/t = 0.125$), but the bands still touch at the three-dimensional Dirac point X . **(c)** A modulation of overlap parameters ($\delta t_1/t = 0.4$) shifts the bulk Dirac point away from the high-symmetry point X , but the gap closes at another point \mathbf{k}^* in the Brillouin zone which is determined by the relation $|\gamma(\mathbf{k}^*)| \doteq 0$ (not shown here). **(d)** Breaking inversion symmetry by adding a staggered sublattice potential $\Delta/t = 0.5$ leads to a true band gap $\mathcal{E}_{\text{gap}} = 2\Delta$ in the absence of spin-orbit coupling. **(e)** The gap induced by the spin-orbit coupling, however, competes with the gap due to the staggered sublattice potential, lifting the degeneracy of both conduction and valence bands and shifting the bulk Dirac point on the high-symmetry line between X and W for $\delta t_j/t = 0$. **(f)** In presence of the sublattice potential, there is a critical spin-orbit coupling $\lambda_{\text{SO}}^c = \Delta/8$, where a quantum phase transition to a trivial band insulator phase for $\lambda_{\text{SO}} < \lambda_{\text{SO}}^c$ takes place.

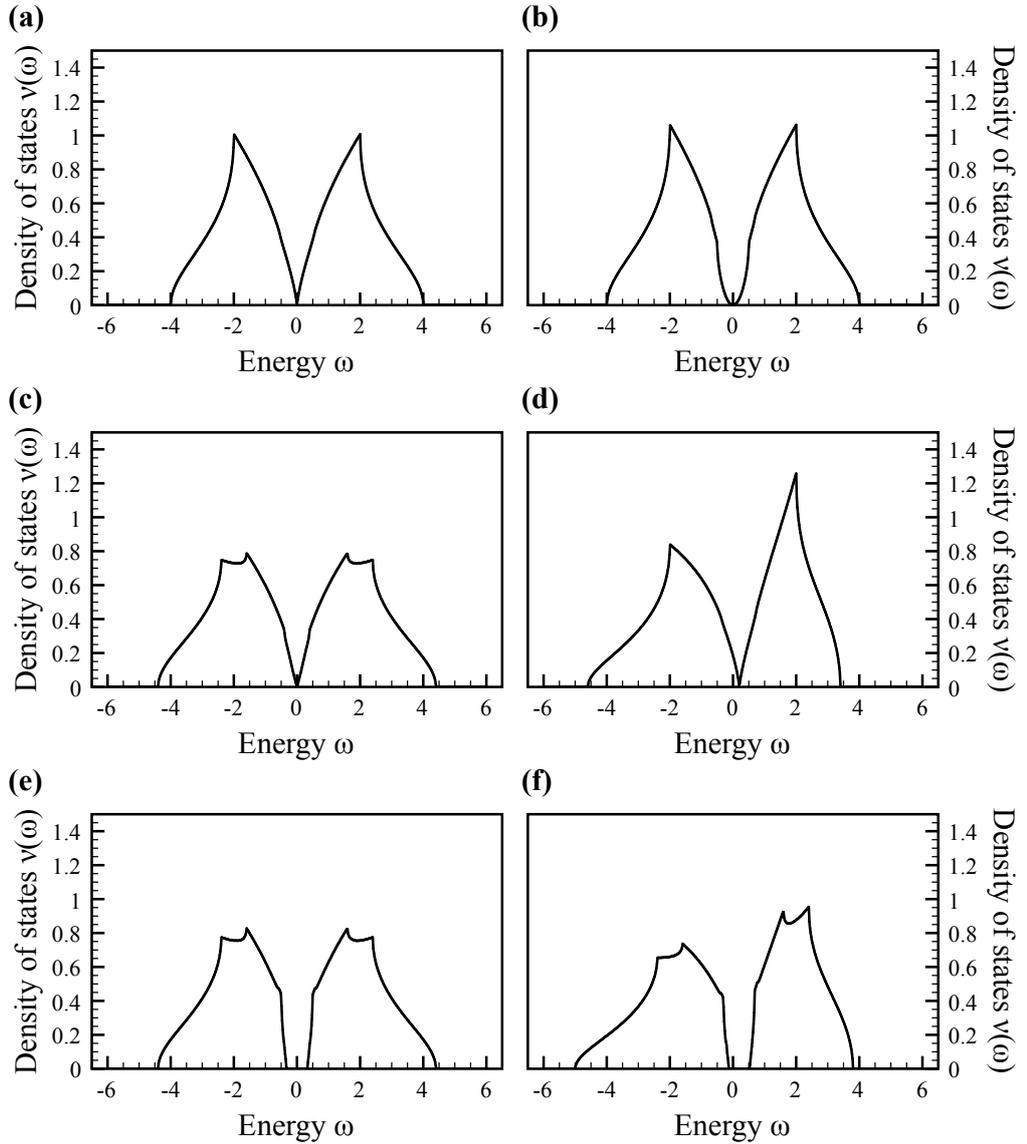


Figure 5.13. Plots of the density of states $\nu(\omega)$ of the Fu-Kane-Mele Hamiltonian (5.43) as function of the energy ω/t . **(a)** For $\lambda_{\text{SO}}/t = \delta t_j/t = t'/t = 0$ both conduction and valence bands are doubly degenerate, and there is a critical high-symmetry line connecting X and W , leading to the soft gap behavior of $\nu(\omega)$ at the Fermi level, $\omega/t = 0$. **(b)** By including a finite spin-orbit coupling parameter ($\lambda_{\text{SO}}/t = 0.125$) this degeneracy of valence and conduction bands along a line is lifted, but the bands still touch at the three-dimensional Dirac point X . As a consequence, $\nu(\omega)$ vanishes quadratically at the Fermi level. **(c)** The modulation of overlap parameters ($\delta t_1/t = 0.4$) only moves the line of bulk Dirac points away from the high-symmetry line connecting X and W , but does not collapse the line to single bulk Dirac points. **(d)** The conduction and valence bands are no longer particle-hole symmetric once a finite next-nearest neighbor interaction $t'/t = 0.05$ is taken into account. This, however, does not remove the line of bulk Dirac points. **(e)** Combining the spin-orbit interaction ($\lambda_{\text{SO}}/t = 0.125$) and the anisotropy hopping along of the bonds ($\delta t_1/t = 0.4$) destroys the line of bulk Dirac points and creates a bulk band gap in the system which finally allows for the existence of gapless surface states. **(f)** The bulk band gap prevails when taking into account the additional next-nearest neighbor interaction $t'/t = 0.05$, but the properties of the gapless surface states can be modified in a non-trivial way by next-nearest neighbor interactions.

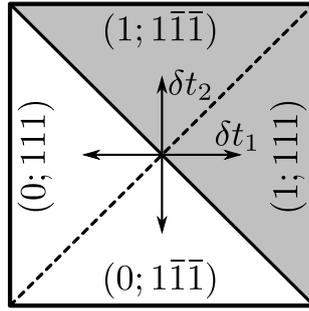


Figure 5.14. Phase diagram as function of the modulation δt_1 and δt_2 of the nearest-neighbor overlap parameters t for bonds in the $[111]$ and $[1\bar{1}\bar{1}]$ directions ($\delta t_{1,2}/t \ll 1$). The gray (white) area indicates the strong (weak) topological insulator phase, and for each phase the topological \mathbb{Z}_2 invariants $(\nu_0; \nu_1\nu_2\nu_3)$ are explicitly given. Note that the indices $(\nu_1\nu_2\nu_3)$ can be interpreted as Miller indices for planes of 2D quantum spin Hall insulators stacked and coupled along the $[\nu_1\nu_2\nu_3]$ direction. Figure after Ref. 8.

metallic surface which is topologically protected by time-reversal symmetry.

Note, however, that transitions between any two phases of the Fu-Kane-Mele model can only occur when the bulk band gap vanishes according to the bulk-boundary correspondence. As shown by Fu, Kane, and Mele, such quantum phase transitions can only occur when the bulk band gap closes at any of the X points in the bulk Brillouin zone of the diamond lattice⁸. Hence, $\delta t_i/t = 0$ ($i = 1, \dots, 4$) can be considered as a multi-critical point in the phase diagram of the Fu-Kane-Mele model (cf. Fig. 5.14) which separates eight different phases:

WTI	(0; 111)	(0; $1\bar{1}\bar{1}$)	(0; $\bar{1}1\bar{1}$)	(0; $\bar{1}\bar{1}1$)
STI	(1; 111)	(1; $1\bar{1}\bar{1}$)	(1; $\bar{1}1\bar{1}$)	(1; $\bar{1}\bar{1}1$)

Finally, note that the Fu-Kane-Mele model does not show a trivial band insulator phase in the perturbatively accessible regime around the multi-critical point. A trivial phase can only be reached when drastically increasing the overlap parameter along one of the bonds, *i.e.*, $\delta t_1/t > 2$, so that electronic interactions along that bond dominate, but we will not consider that phase.

5.3.3 Surface bands in diamond slabs

So far we have only discussed the bulk phase diagram of the Fu-Kane-Mele model in terms of the electronic band structure $\mathcal{E}_{\pm\pm}(\mathbf{k})$ and the corresponding density of states $\nu(\omega)$. As shown by Fu, Kane, and Mele this model has also non-trivial surface states which can be studied by solving the tight-binding Fu-Kane-Mele Hamiltonian in a finite slab geometry with open boundary conditions in the growth direction of the slab and periodic boundary conditions in the two perpendicular directions. Figs. 5.15 and 5.16 show the two-dimensional slab band structures for two slab geometries with (001) and (111) surfaces along the high-symmetry points and lines of the corresponding surface Brillouin zone. Note that Fig. 3.9 schematically illustrates the construction of the surface Brillouin zone for different normal vectors in the $[001]$, $[011]$, and $[111]$ directions, and this construction is valid for both zinc-blende and diamond lattice as they have the same reciprocal lattice.

As expected, in the strong topological insulator phases we observe an odd number of Dirac points in the surface spectrum indicated by red solid lines in Fig. 5.15, while the weak topological insulator phase is characterized by an even number of Dirac points, as shown in Fig. 5.16. Those states crossing the bulk band gap are exponentially localized at the surfaces, giving rise

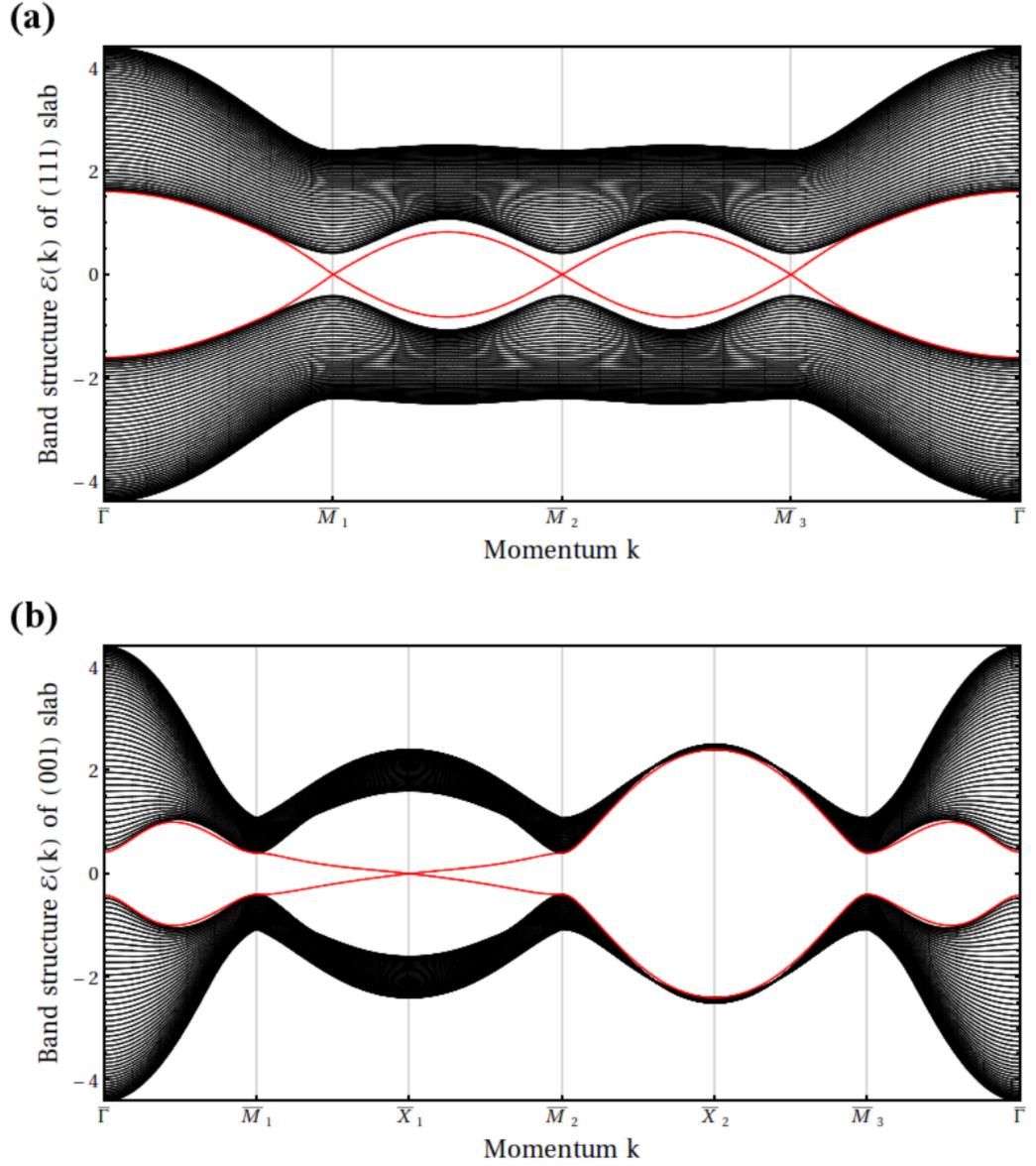


Figure 5.15. Electronic band structure of slabs with (a) (111) surfaces and (b) (001) surfaces for a slab width of $N = 50$ layers. The parameters of the model are chosen as $t = 1$, $\delta t_1/t = 0.4$, and $\lambda_{\text{SO}}/t = 0.125$, so that both panels realize a strong topological insulator. In addition to the bulk bands, we observe surface bands traversing the bulk band gap indicated by red lines, and those are localized at the top and bottom surfaces of the slab. In general, in the strong topological insulator phase there exists an odd number of points, where the surface bands cross in a Dirac-like fashion, as can be seen in both panels. On the (111) surfaces, this can either happen at the three non-equivalent \bar{M}_1 , \bar{M}_2 , and \bar{M}_3 points or at the $\bar{\Gamma}$ point, and which one of the two cases is realized depends on the surface termination, as discussed in the main text. Similarly, we find a single Dirac point on the (001) surfaces either at the \bar{X}_1 point or at the \bar{X}_2 point.

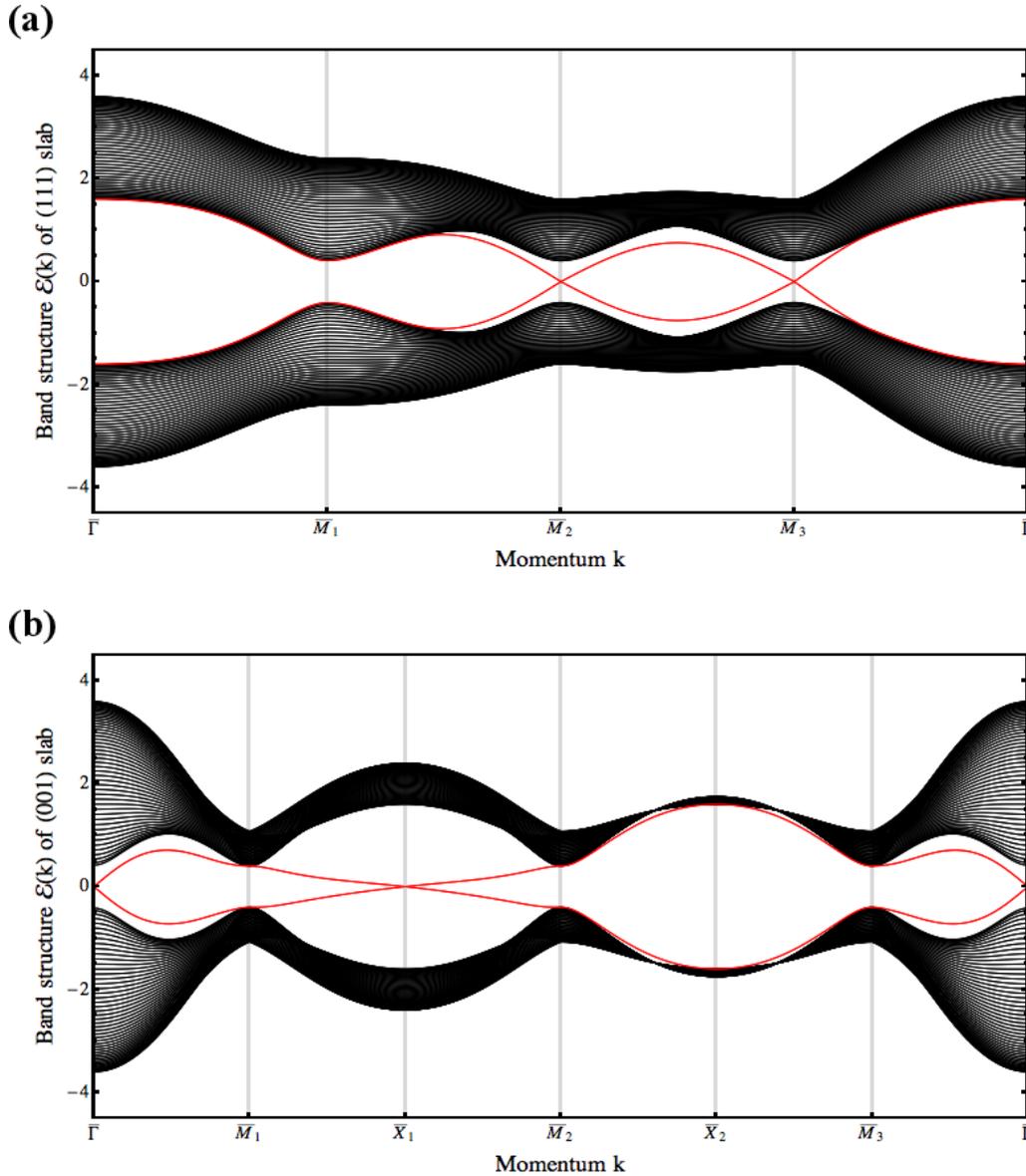


Figure 5.16. Electronic band structure of slabs with (a) (111) surfaces and (b) (001) surfaces for a slab width of $N = 50$ layers. The parameters of the model are chosen such that both panels realize the weak topological insulator with an even number of Dirac cones on the surfaces for $t = 1$ and $\lambda_{SO}/t = 0.125$. In panel (a) we have chosen $\delta t_2/t = -0.4$, while in panel (b) we use $\delta t_1/t = 0.4$. Obviously, there exist surface bands which traverse the bulk band gap, but in contrast to the strong topological insulator shown in Fig. 5.15 those surface bands always cross in an even number of Dirac points.

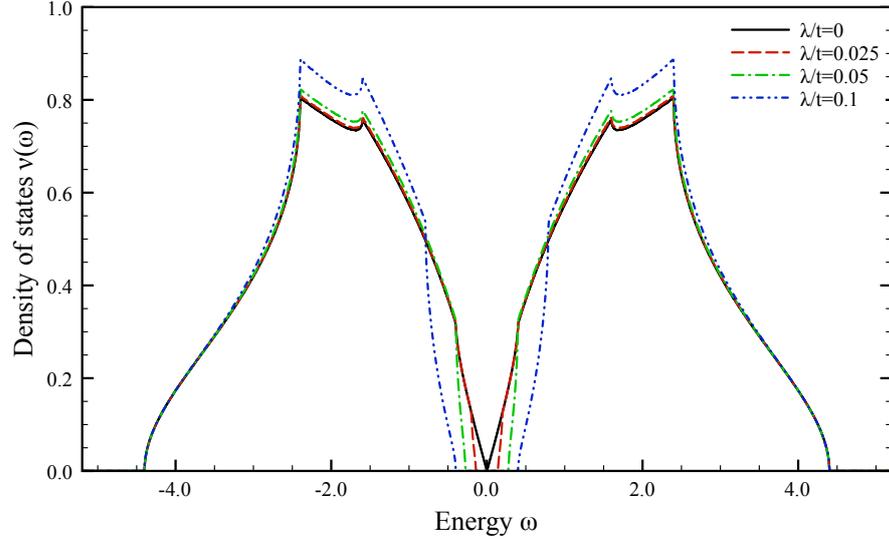


Figure 5.17. Density of states $\nu(\omega)$ of the Fu-Kane-Mele model as function of the energy ω/t for different values of the spin-orbit coupling λ_{SO}/t . The parameters of the model are chosen such that the system is in a strong topological insulator phase for $\lambda_{\text{SO}}/t > 0$: $t = 1$, $\delta t_1/t = 0.4$, and $\Delta/t = 0$. For small spin-orbit interactions λ_{SO} the bulk band gap $\mathcal{E}_{\text{gap}} \propto \lambda_{\text{SO}}$ and vanishes at the quantum critical point $\lambda_{\text{SO}}^c/t = 0$ (indicated as black line).

to the notion of a 2D surface metal of Dirac fermions in 3D topological insulators. Due to the Kramers' theorem, the Dirac points are located at the **time-reversal invariant momenta** of the surface Brillouin zones. Note that in the strong topological phases of the Fu-Kane-Mele model, each surface state traversing the bulk band gap has a single time-reversed partner state, and those pairs of eigenstates are robust against small perturbations due to time-reversal symmetry.

It is interesting to study the effect of the different model parameters on the Fermi velocity v_F^{surf} of the surface Dirac fermions. To be specific, we focus on the strong topological insulator phase of a slab with (001) surfaces in which the bond along the [111] direction is stronger than the others, *i.e.*, we consider only $\delta t_1/t = 0.4$, but the results presented in the following are equally valid in the other phases with $\nu_0 = 1$ as well.

Let us first discuss the role of the spin-orbit coupling parameter λ_{SO} . In an inversion-symmetric system, *i.e.*, for $\Delta/t = 0$, the spin-orbit interaction between next-nearest neighboring atoms leads to a finite bulk band gap which depends linearly on the strength of the spin-orbit coupling λ_{SO} for small couplings, *i.e.*, $\mathcal{E}_{\text{gap}} \propto \lambda_{\text{SO}}$. This can be easily seen from the density of states shown in Fig. 5.17 for different spin-orbit couplings $\lambda_{\text{SO}}/t \in \{0.1, 0.05, 0.025, 0\}$. Apparently, the bulk band gap vanishes at the quantum critical point which is given by $\lambda_{\text{SO}}^c/t = 0$ in an inversion-symmetric system ($\Delta/t = 0$). Note that only at that point the system may undergo a quantum phase transition from a topological quantum state of matter to another one, as mentioned before.

Furthermore, since the bulk band gap closes at the bulk quantum critical point $\lambda_{\text{SO}}^c/t = 0$, the surface states traversing the bulk band gap are expected to become perfectly flat in that limit. In other words, the surface Fermi velocity v_F^{surf} is expected to vanish as λ_{SO} decreases:

$$v_F^{\text{surf}} \propto \lambda_{\text{SO}} \rightarrow 0 \quad \text{as} \quad \lambda_{\text{SO}} \rightarrow 0. \quad (5.51)$$

Fig. 5.18 (a–d) shows the expected evolution of the surface states obtained from a (001) slab

as function of decreasing spin-orbit coupling $\lambda_{\text{SO}}/t \in \{0.125, 0.05, 0.025, 0\}$. In principle, the decrease of the surface Fermi velocity opens up the possibility to observe a spontaneous mass generation on the surfaces of the topological insulator since $\alpha = e^2/(\hbar v_F^{\text{surf}})$ could become large, but as we show below, the dielectric constant ϵ also diverges and α goes to zero instead of diverging. Nevertheless, when the average kinetic energy goes to zero, even local, Hubbard-like interactions become more relevant and can drive the transition, as we further discuss below.

Let us now discuss the role of the hopping term t' between next-nearest neighboring atoms as leading-order correction to the usual hopping term t . It is well known from graphene that next-nearest neighbor interactions shift the Fermi level and change the bandwidths of both valence and conduction bands, thus breaking particle-hole symmetry. In addition, the edge states in zigzag graphene nanoribbons become dispersive as well, with the bandwidth set by the next-nearest neighbor interaction t' (Ref. 155). In the Fu-Kane-Mele model on the diamond lattice, the surface bands also become “flat” at the Dirac point located at \bar{X} in the sense that $\partial_{\mathbf{k}}\mathcal{E}(\mathbf{k}) = 0$ for $\lambda_{\text{SO}} \rightarrow 0$, as shown in Fig. 5.18 (e–h). However, due to next-nearest neighbor interactions, away from the \bar{X} point the surface bands also become dispersive and bend up- or downwards, again with a bandwidth set by t' . As a consequence, in the presence of interactions, the surface bands will couple to the bulk bands due to the finite curvature, and this allows for elastic scattering into the bulk. The opening of an elastic scattering channel for the Dirac fermions occurs at a finite spin-orbit interaction strength which is set by the next-nearest neighbor hopping amplitude, $\lambda_{\text{SO}} \approx t'/4$. Hence, to leading order the decrease of the surface Fermi velocity v_F^{surf} as function of the spin-orbit coupling λ_{SO} is cut off by a finite next-nearest neighbor hopping t' , resulting in the rough estimate $v_F^{\text{surf}} \propto \max\{|\lambda_{\text{SO}}|, t'\}$. In a sense, the magnitude of the surface Fermi velocity v_F^{surf} is reduced by a factor t'/t compared to the bulk Fermi velocity, *i.e.*, $v_F^{\text{surf}} = (t'/t)v_F^{\text{bulk}}$ for $\lambda_{\text{SO}}/t < t'/t$. Furthermore, note that the average kinetic energy of the Dirac fermions does not vanish as the surface bands now have a finite bandwidth.

Finally, let us consider the effect of a staggered sublattice potential Δ which breaks the inversion symmetry of the diamond lattice, thereby reducing the point group symmetry of the crystal lattice from O_h^7 to T_d^2 (cf. sections 3.3.1 and 5.3.1 for details on the crystal symmetries of the zinc-blende and diamond lattices). Most importantly, the bulk quantum critical point in the phase diagram is shifted towards a finite spin-orbit coupling $\lambda_{\text{SO}}^c = \Delta/8$, where the bulk band gap between the $\mathcal{E}_{\pm,-}(\mathbf{k})$ bands vanishes [cf. Eq. (5.49)]. Fig. 5.19 shows the density of states of the inversion-asymmetric Fu-Kane-Mele model for different spin-orbit couplings. It turns out that for $\lambda_{\text{SO}} > \lambda_{\text{SO}}^c$ the Fu-Kane-Mele model describes a strong topological insulator, while for $\lambda_{\text{SO}} < \lambda_{\text{SO}}^c$ the system is a trivial band insulator. This situation is reminiscent of the phase diagram of the Kane-Mele model on the 2D graphene lattice, where the quantum spin Hall insulator phase is separated from a trivial band insulator state by a quantum critical point at $\lambda_{\text{SO}}^c = 3\sqrt{3}\Delta$ (Refs. 3,4).

Fig. 5.18 (i–l) shows a sequence of slab band structures in the presence of both inversion asymmetry ($\Delta/t = 0.25$) and next-nearest neighbor hopping ($t'/t = 0.1$) as function of the spin-orbit coupling $\lambda_{\text{SO}}/t \in \{0.125, 0.05, 0.025, 0\}$. Above the critical spin-orbit interaction strength $\lambda_{\text{SO}}^c/t = \Delta/(8t) = 0.03125$ we observe the expected surface bands which are split due to inversion asymmetry [panels (i, j)]. Note, however, that one of the surface states shows a Dirac point with an energy that lies in the projected bulk valence bands, and thus the corresponding surface state is likely to interact with the bulk bands due to elastic scattering. When the bulk valence and conduction bands touch at the bulk quantum critical point $\lambda_{\text{SO}}^c/t = 0.03125$, the surface states can “unknot” and remove their topologically non-trivial winding over the Brillouin zone. As a consequence, for $\lambda_{\text{SO}} < \lambda_{\text{SO}}^c$ the Fu-Kane-Mele model describes

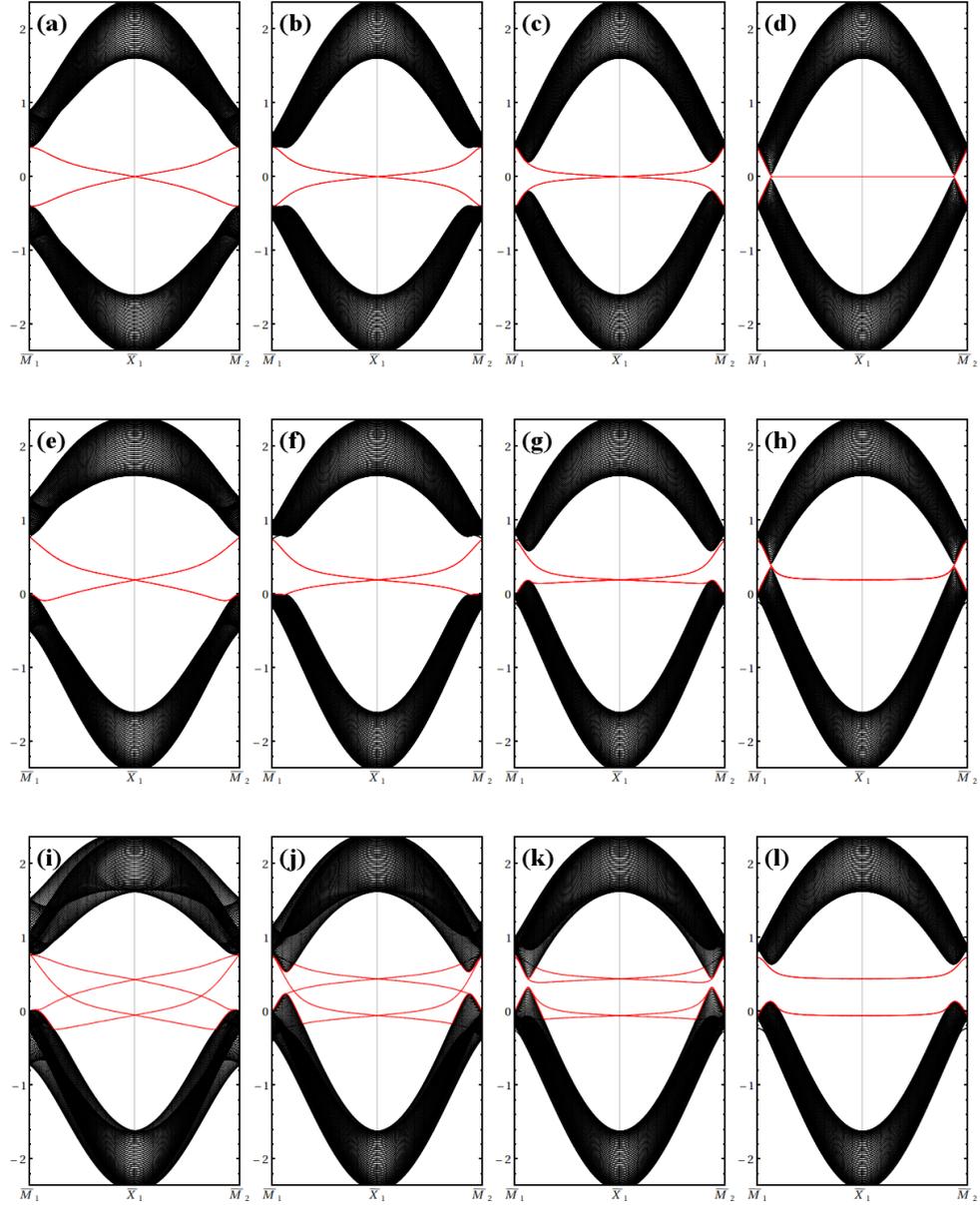


Figure 5.18. Plots of the electronic band structure for a slab with (001) surface and a thickness of $N = 50$ layers. The bulk bands are indicated by black lines, while the surface states traversing the bulk band gap are shown in red. The parameters of the model are chosen such that it realizes a strong topological insulator in the simplest case: $t = 1$, $\delta t_1/t = 0.4$, $\lambda_{\text{SO}}/t = 0.125$, $t'/t = 0$, and $\Delta/t = 0$ [cf. panel (a)]. Columns from left to right show band structure plots for $\lambda_{\text{SO}}/t \in \{0.125, 0.05, 0.025, 0\}$, while the rows from top to bottom show plots for $(t'/t, \Delta/t) = (0, 0)$, $(0.1, 0)$, and $(0.1, 0.25)$, respectively. (a–d) In the absence of next-nearest neighbor hopping ($t'/t = 0$) and inversion asymmetry ($\Delta/t = 0$) the surface bands are doubly degenerate, so that top and bottom surfaces exhibit identical surface states. Those surface states become perfectly flat in the limit of vanishing spin-orbit interactions, i.e., $v_F^{\text{surf}} \rightarrow 0$ for $\lambda_{\text{SO}} \rightarrow 0$. (e–h) Taking next-nearest neighbor interactions into account ($t'/t = 0.1$) the surface states do not become perfectly flat in the limit $\lambda_{\text{SO}} \rightarrow 0$, but exhibit a finite band curvature, where the bandwidth of the surface band within the bulk band gap is set by the nearest-neighbor hopping amplitude t' . (i–l) In the absence of inversion symmetry, $\Delta/t = 0.25$, the degeneracy of the surface bands is lifted, but the states corresponding to top and bottom surfaces still cross in Dirac points located at the \bar{X} point of the surface Brillouin zone. There exists, however, a critical spin-orbit coupling $\lambda_{\text{SO}}^c = \Delta/8$, where the bulk band gap closes and the system undergoes a quantum phase transition to a trivial band insulator phase [in between panels (j) and (k)].

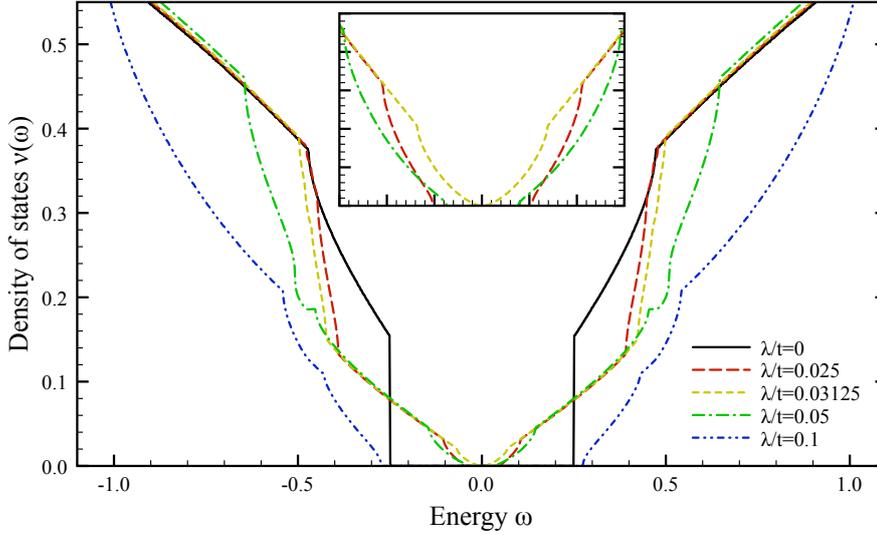


Figure 5.19. Density of states $\nu(\omega)$ of the Fu-Kane-Mele model with broken inversion symmetry as function of the energy ω/t for different values of the spin-orbit coupling λ_{SO}/t . The parameters of the model ($t = 1$, $\delta t_1/t = 0.4$, and $\Delta/t = 0.25$) are chosen such that the system is in a strong topological insulator phase for $\lambda_{\text{SO}} > \lambda_{\text{SO}}^c = \Delta/8$. The inset shows that the bulk band gap closes at the critical spin-orbit coupling $\lambda_{\text{SO}}^c/t = \Delta/(8t) = 0.03125$ and the corresponding density of states vanishes quadratically in agreement with isolated Dirac points (indicated by the yellow dashed line). Although on both sides of the quantum phase transition the system remains gapped, the profile of the density of states $\nu(\omega)$ as function of ω/t is quite different, showing a discontinuous vanishing in the trivial phase.

a trivial band insulator phase. Similar to the previous discussions we observe that the surface bands become increasingly flat as we approach the bulk quantum critical point λ_{SO}^c , which might allow for elastic scattering between the surface bands and the bulk states [see, e.g., Fig. 5.18 (i)].

5.3.4 Topological origin of the flat surface bands

As shown in the previous section, solving the Fu-Kane-Mele model in a slab geometry we find surface states which become perfectly flat in the limit $\lambda_{\text{SO}} \rightarrow 0$, but this property is not limited to a specific model or geometry, as we discuss below. More generally, in the presence of a chiral symmetry, the system supports nodal lines, *i.e.*, lines with zero (or constant) energy in the bulk, and their existence leads to flat surface bands, *i.e.*, topologically protected gapless fermions localized at the surfaces forming a band without dispersion^{176–179}.

To gain a better understanding of those flat bands, we consider a semimetal with a nodal line in the bulk. The topological invariant supporting the existence and topological stability of the nodal line is the following contour integral^{176,177}:

$$\nu = \frac{1}{N} \oint_{\mathcal{C}} \frac{d\mathbf{k}}{2\pi i} \cdot \text{tr} \{ \hat{\Sigma} \mathcal{H}(\mathbf{k})^{-1} \nabla_{\mathbf{k}} \mathcal{H}(\mathbf{k}) \} \in \mathbb{Z}, \quad (5.52)$$

where N denotes the number of conduction and valence bands, and $\hat{\Sigma}$ defines the **chiral symmetry** operator. The topological protection stems from the fact that the phase of the zero-frequency Green's function $\mathcal{G}(\omega = 0, \mathbf{k}) = \mathcal{H}(\mathbf{k})^{-1}$ can only change by an integer multiple of 2π when going around the nodal line in the three-dimensional momentum space (see Fig. 5.20).

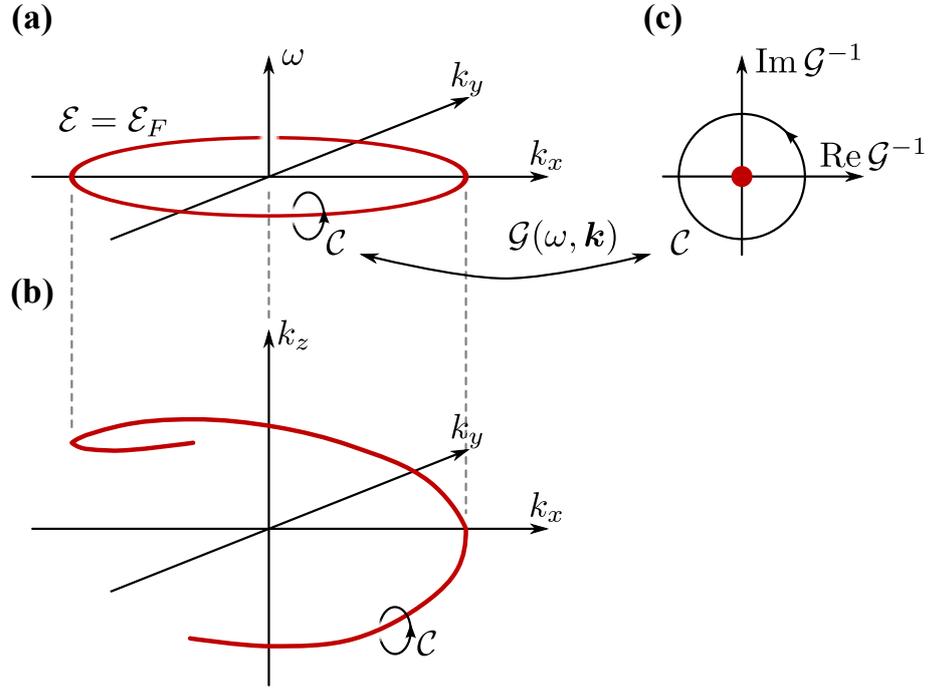


Figure 5.20. **a, b)** The Fermi surface $\mathcal{E}(\mathbf{k}) = \mathcal{E}_F$ indicated by the red line is the momentum-space analogue of a vortex line, where the phase of the Green's function changes by an integer multiple of 2π going around the nodal line in (ω, \mathbf{k}) space. This nodal has a non-zero winding number in momentum space, and this gives rise to topologically protected states located on the surfaces of the system. Those surface states form a completely flat band which terminates on the projection of the nodal line onto the surface. **c)** Winding of the Green's function \mathcal{G} around the nodal line which can be considered as a singularity of the Green's function (red dot). Figure after Ref. 176.

On the other hand, as we are concerned with certain slab geometries, one may also choose the contour as a straight line $[-\pi/a, \pi/a]$ along the direction k_\perp normal to the surface, and consider the in-plane momentum \mathbf{k}_\parallel in the surface Brillouin zone as a parameter of the winding number:

$$\nu(\mathbf{k}_\parallel) = \frac{1}{N} \int_{-\pi/a}^{\pi/a} \frac{dk_\perp}{2\pi i} \text{tr} \{ \hat{\Sigma} \mathcal{H}(k_\perp, \mathbf{k}_\parallel)^{-1} \partial_{k_\perp} \mathcal{H}(k_\perp, \mathbf{k}_\parallel) \} \in \mathbb{Z}. \quad (5.53)$$

Since the points $k_z = \pm\pi/a$ are equivalent due to periodic boundary conditions, the contour $[-\pi/a, \pi/a]$ forms a closed loop as well, and thus the above integral takes on integer values if the integration path does not cross the bulk nodal line, where the energy is zero.

Now let us assume that for any point \mathbf{k}_\parallel within the projection of the nodal line onto the surface $\nu(\mathbf{k}_\parallel) \neq 0$, while $\nu(\mathbf{k}_\parallel) = 0$ outside this region [cf. Fig. 5.20 (a)]. Since the states with non-trivial winding number $\nu(\mathbf{k}_\parallel) \neq 0$ cannot be adiabatically transformed into the states of topologically trivial systems with $\nu(\mathbf{k}_\parallel) = 0$ like the vacuum, a surface state with zero energy emerges for each momentum \mathbf{k}_\parallel inside the non-trivial region of the surface Brillouin zone at the interface between systems with different winding numbers. Note that the appearance of topologically protected surface states has the same mathematical footing as in the Kane-Mele model discussed in section 5.2.6 and in the Su-Schrieffer-Heeger model discussed in appendix C—namely the fact that the first homotopy group of the unit circle S^1 is non-trivial: $\pi_1(S^1) = \mathbb{Z}$. Furthermore, following the arguments presented in those sections, it also becomes immediately clear that the existence of flat surface bands in the Fu-Kane-Mele model is

protected by the chiral symmetry $\hat{\Sigma}$.

Now let us discuss the flat surface bands in the Fu-Kane-Mele model in the absence of spin-orbit interactions λ_{SO} , modulations of the overlap parameters δt_j , hopping between next-nearest neighbors t' , and the staggered sublattice potential Δ . The remaining nearest-neighbor hopping couples the \mathfrak{A} and \mathfrak{B} sublattice of the diamond lattice, and the Bloch Hamiltonian has a chiral symmetry which is given by the operator $\hat{\Sigma} \equiv \tau_z \otimes \sigma_0$. Moreover, the Fu-Kane-Mele model has nodal lines in the bulk connecting the high-symmetry points X and W in the Brillouin zone for $\delta t_i/t = 0$ with $i = 1, \dots, 4$ [cf. Fig. 5.12 (a)]. Those nodal lines are shifted in the bulk Brillouin zone if we consider finite modulations δt_j of the hopping amplitude, but their presence still guarantees the existence of topologically protected surface states. We have calculated the winding number $\nu(\mathbf{k}_{\parallel})$ [cf. Eq. (5.53)] in that situation for slabs with (001) and (111) oriented surfaces, as shown in Fig. 5.21. As discussed above, regions with non-trivial winding number $\nu(\mathbf{k}_{\parallel})$, which cannot be connected adiabatically to the vacuum, host perfectly flat surface bands. Indeed, for the (001) slabs we find such zero-energy states, shown in Fig. 5.18 (d), in the same area of the surface Brillouin zone, where $\nu(\mathbf{k}_{\parallel}) = 1$, as shown in Fig. 5.21 (a). Similarly, we also find flat surface bands in the (111) oriented slabs which are in perfect agreement with Fig. 5.15 when we take the limit $\lambda_{\text{SO}}/t = 0$.

However, note that in all slab geometries, the winding number $\nu(\mathbf{k}_{\parallel})$ depends on the termination of the surfaces, which can be understood as follows: As shown in appendix C, the winding number $\nu(\mathbf{k}_{\parallel})$ can also be expressed in terms of a unit vector $\hat{\mathbf{d}}(\mathbf{k})$, where $\mathbf{d}(\mathbf{k}) \in \mathbb{R}^5$ is defined by the parametrization of the bulk Bloch Hamiltonian as $\mathcal{H}(\mathbf{k}) = \mathbf{d}(\mathbf{k}) \cdot \mathbf{\Gamma}$. Here, $\mathbf{\Gamma}$ is a five-component vector of the 4×4 Γ matrices similar to the ones introduced in chapter 3 for our minimal model. Aside from the number and position of the nearest-neighbors in the crystal lattice, also the choice of the unit cell influences the vector $\mathbf{d}(\mathbf{k})$. In the simplest case, the unit cell of the diamond lattice contains only two sublattices \mathfrak{A} and \mathfrak{B} , which allows for different definitions of $\mathbf{d}(\mathbf{k})$, for example:

$$\mathbf{d}_{\mathfrak{A}\mathfrak{B}}(\mathbf{k}) \equiv -t \sum_{j=1}^4 e^{i\mathbf{k} \cdot (\mathbf{d}_j - \mathbf{d}_1)}, \quad (5.54a)$$

$$\mathbf{d}_{\mathfrak{B}\mathfrak{A}}(\mathbf{k}) \equiv -t \sum_{j=1}^4 e^{i\mathbf{k} \cdot (\mathbf{d}_j - \mathbf{d}_2)}. \quad (5.54b)$$

Since $\mathbf{d}_{\mathfrak{A}\mathfrak{B}}(\mathbf{k}) = \exp[i\mathbf{k} \cdot (\mathbf{d}_2 - \mathbf{d}_1)] \mathbf{d}_{\mathfrak{B}\mathfrak{A}}(\mathbf{k})$ and $\mathbf{d}_2 - \mathbf{d}_1 = \frac{a}{2}(0, -1, -1)^T$, this results in a change of the winding number by $\Delta\nu = 1$, as shown in panels (b) and (c) of Fig. 5.21 for different terminations of (111) slabs.

In the presence of perturbations breaking chiral symmetry, however, those surface bands become dispersive with a bandwidth set by the perturbation strength. For example, Fig. 5.18 (a–c) shows dispersive surface bands with a bandwidth proportional to the spin-orbit coupling λ_{SO} , because spin-orbit interactions are proportional to $\tau_z \otimes \sigma_z$, breaking chiral symmetry as $\{\hat{\Sigma}, \tau_z \otimes \sigma_z\} \neq 0$. We have also checked by explicit tight-binding calculations that hopping between second-nearest neighbors t' , which also breaks the chiral symmetry of the Fu-Kane-Mele model, leads to a finite dispersion of the surface bands [cf. Fig. 5.18 (d)].

Finally, an important consequence of a flat surface band is the singular density of states which may lead to instabilities of the surface Dirac metal towards the formation of symmetry-broken states with smaller density of states. For example, Heikkilä and Volovik have shown that topologically protected surface bands without dispersion emerge by stacking graphene layers on top each other in a rhombohedral stacking^{178,179}. The flat surface bands arise, because a

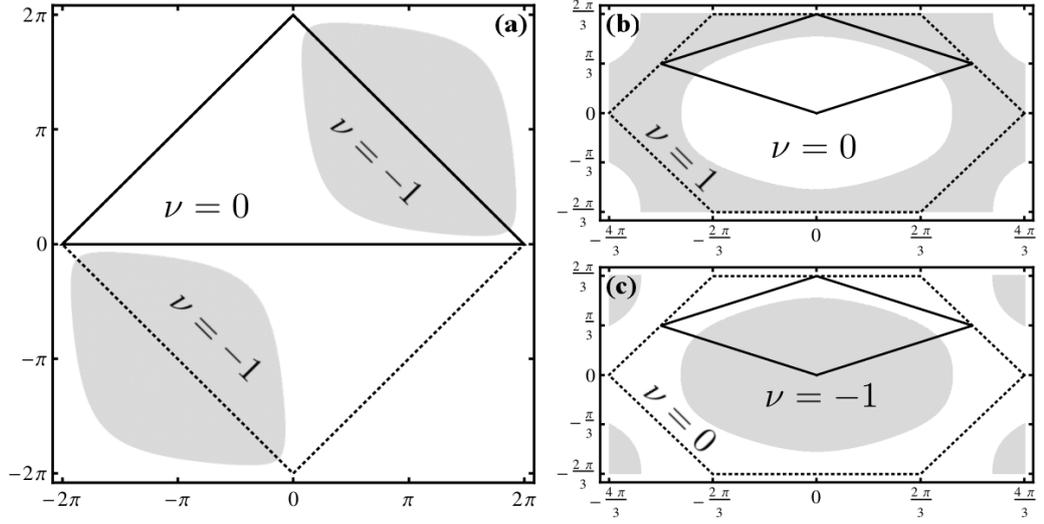


Figure 5.21. Plots of the winding number $\nu(\mathbf{k}_{\parallel})$ as function of the momentum \mathbf{k}_{\parallel} in the surface Brillouin zone of (a) (001) surfaces and (b, c) (111) surfaces, where gray and white areas correspond to winding numbers $\nu(\mathbf{k}_{\parallel}) = \pm 1$ and $\nu(\mathbf{k}_{\parallel}) = 0$, respectively. In all plots, the surface Brillouin zone is indicated by dotted lines, and the path through the surface Brillouin zone used in Figs. 5.15 and 5.16 is marked by solid lines. In panels (b, c) we used different terminating layers of the (111) slabs which can be taken into account by a redefinition of the Bloch Hamiltonian $\mathcal{H}(\mathbf{k})$, thereby resulting in different winding numbers.

nodal line is formed in the bulk of the system in the limit of a large number of graphene layers. Furthermore, in that model the formation of a singular density of states with increasing number of layers N is exemplified by the singularity $\nu(\omega) \sim \omega^{2/N-1}$ which for $N \gg 1$ yields a $1/\omega$ divergence: $\nu(\omega) = 1/(2\pi N\omega)^{178,179}$. So far we have shown that in the Fu-Kane-Mele model there exist flat surface bands which, in principle, opens up the possibility for a spontaneously generated mass in the presence of long-range electron-electron interactions. However, the flat band also leads to a sharp peak in the density of states, and therefore we have to discuss the polarization function of the Fu-Kane-Mele model in the limit $\lambda_{SO}/t \rightarrow 0$ as well.

5.3.5 Polarization function for the Fu-Kane-Mele model

So far we have only discussed the evolution of the Fermi velocity v_F^{surf} upon tuning the strength of the spin-orbit interactions $\lambda_{SO}/t \rightarrow 0$, where we have already seen that nodal lines appear in the bulk of the system at the quantum critical point $\lambda_{SO}/t = 0$. Since the density of states vanishes at the Fermi level, *i.e.*, $\nu(\omega \rightarrow 0) = 0$, the bulk is not metallic, and naively one would not expect screening to be relevant. Consequently, the bulk dielectric should be rather constant so that the effective interaction strength α diverges. However, our calculations show that this is not the fact, and we find a precursor to screening in the divergence of the polarization function in the low-frequency limit $\omega = 0$ and long-wavelength limit $\mathbf{q} \rightarrow 0$. Here, we consider the dielectric constant $\epsilon(\lambda_{SO})$ as function of the spin-orbit coupling λ_{SO} which can be easily calculated from the electronic polarization function $\Pi(\omega, \mathbf{q})$ in the static limit.

Our starting point for computing the polarization function $\Pi(\omega, \mathbf{q})$ is the functional representation of the quantum partition function \mathcal{Z} expressed as a coherent state path integral over the action \mathcal{S} :

$$\mathcal{Z} = \int \mathcal{D}(\Psi^\dagger, \Psi) \exp(-\mathcal{S}[\Psi^\dagger, \Psi]) \quad \text{with} \quad \mathcal{S}[\Psi^\dagger, \Psi] = \mathcal{S}_0[\Psi^\dagger, \Psi] + \mathcal{S}_{\text{int}}[\Psi^\dagger, \Psi]. \quad (5.55)$$

The **free action** $\mathcal{S}_0[\Psi^\dagger, \Psi]$ describing non-interacting spinful electrons in terms of the Fu-Kane-Mele Hamiltonian $\mathcal{H}_{\text{FKM}}(\mathbf{k})$ [cf. Eq. (5.43)] is simply given by

$$\mathcal{S}_0[\Psi^\dagger, \Psi] \equiv \int \frac{d^4 k}{(2\pi)^4} \Psi_k^\dagger (-i\nu \mathbb{1} + \mathcal{H}_{\text{FKM}}(\mathbf{k})) \Psi_k, \quad (5.56)$$

where, for brevity, we have introduced the “4-momentum” $k \equiv (\nu, \mathbf{k})$ comprising both frequency ν and momentum \mathbf{k} , and Ψ_k denotes the four-component spinors with sublattice and spin degrees of freedom. Furthermore, we consider an instantaneous Coulomb interaction between the Dirac fermions as their Fermi velocity is much smaller than the speed of light:

$$\mathcal{S}_{\text{int}}[\Psi^\dagger, \Psi] = \frac{1}{2} \int d^4 k \int d^4 k' \int \frac{d^4 q}{(2\pi)^4} \Psi_{k+q}^\dagger \Psi_{k'-q}^\dagger V(\mathbf{q}) \Psi_{k'} \Psi_k, \quad (5.57)$$

where $V(\mathbf{q}) = e^2/|\mathbf{q}|^2$ is the Fourier transform of the bare Coulomb potential $V(\mathbf{r}) = e^2/|\mathbf{r}|$. The **free Green’s function** $\mathcal{G}_0(k)$ of the electron gas is defined as the Wick contraction $\langle \Psi_k \Psi_k^\dagger \rangle_0$. In terms of the Fu-Kane-Mele Hamiltonian, $\mathcal{G}_0(k)$ is just the inverse of the kernel of the free action $\mathcal{S}_0[\Psi^\dagger, \Psi]$:

$$\mathcal{G}_0(k) \equiv \langle \Psi_k \Psi_k^\dagger \rangle_0 = (-i\nu \mathbb{1} + \mathcal{H}_{\text{FKM}}(\mathbf{k}))^{-1}. \quad (5.58)$$

In general, the **polarization function** $\Pi(q)$ with 4-momentum q is defined as the Fourier transform of the connected correlation function $\langle \Psi^\dagger(\mathbf{x}, \tau) \Psi(\mathbf{x}, \tau) \Psi^\dagger(\mathbf{x}', \tau') \Psi(\mathbf{x}', \tau') \rangle_0$. In the case of a translationally invariant system, $\Pi(q)$ can be rewritten as a four-dimensional momentum integral (see Fig. 5.22):

$$\Pi(q) \equiv 2 \int \frac{d^4 k}{(2\pi)^4} \text{tr}\{\mathcal{G}_0(k) \mathcal{G}_0(k+q)\}. \quad (5.59)$$

Here, the factor of 2 stems from the Wick contraction of field operators after relabeling momentum indices, and the trace is evaluated over the internal degrees of freedom, *i.e.*, spin and sublattice degrees of freedom in our case. Within the random phase approximation, the dielectric constant ϵ is then related to the polarization function $\Pi(q)$ by:

$$\epsilon(q) = 1 + \frac{4\pi e}{|\mathbf{q}|^2} \Pi(q). \quad (5.60)$$

Hence, to determine the screening properties of the Fu-Kane-Mele model close to the bulk quantum critical point we have to calculate the polarization function (see Fig. 5.22) in the static limit and for long wavelengths, $\Pi(\omega = 0, \mathbf{q} \rightarrow 0)$. Note that as long as $\lambda_{\text{SO}}/t \neq 0$ the bulk remains gapped, so that we may perform a Taylor expansion to quadratic order in small \mathbf{q} . In contrast, at the quantum critical point $\lambda_{\text{SO}}/t = 0$ we have to consider the full expression for the polarization bubble, as derived in the following.

To calculate the polarization function $\Pi(q)$, we start from Eq. (5.43) and rewrite the Fu-Kane-Mele Hamiltonian $\mathcal{H}_{\text{FKM}}(\mathbf{k})$ in terms of 4×4 Γ -matrices Γ_a as

$$\mathcal{H}_{\text{FKM}}(\mathbf{k}) = \mathbf{d}(\mathbf{k}) \cdot \mathbf{\Gamma}, \quad (5.61)$$

where the five-component vectors $\mathbf{d}(\mathbf{k})$ and $\mathbf{\Gamma}$ are defined as follows:

$$\mathbf{d}(\mathbf{k}) \equiv (\text{Re } \gamma(\mathbf{k}), -\text{Im } \gamma(\mathbf{k}), \lambda_{\text{SO}} u_x(\mathbf{k}), \lambda_{\text{SO}} u_y(\mathbf{k}), \lambda_{\text{SO}} u_z(\mathbf{k}))^T, \quad (5.62a)$$

$$\mathbf{\Gamma} \equiv (\tau_x \otimes \sigma_0, \tau_y \otimes \sigma_0, \tau_z \otimes \sigma_x, \tau_z \otimes \sigma_y, \tau_z \otimes \sigma_z)^T. \quad (5.62b)$$

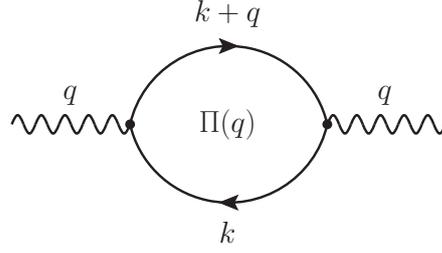


Figure 5.22. Leading-order Feynman diagram for the polarization function $\Pi(q) = \Pi(\omega, \mathbf{q})$. The “4-momenta” k and q comprise both frequency and momentum, *i.e.*, we consider $k \equiv (\nu, \mathbf{k})$ and $q \equiv (\omega, \mathbf{q})$.

Similar to the minimal model introduced in chapter 3, those Γ matrices satisfy the usual Clifford algebra, *i.e.*, they anti-commute with

$$\{\Gamma_a, \Gamma_b\} = 2\delta_{ab} \mathbb{1} \quad \text{for } a, b = 1, \dots, 5. \quad (5.63)$$

Since $\Gamma_a^2 = \mathbb{1}$, we can easily diagonalize the Hamiltonian to obtain the electronic band structure. As a result, we obtain the same result as discussed previously:

$$\mathcal{E}(\mathbf{k}) = \pm |\mathbf{d}(\mathbf{k})|, \quad (5.64)$$

where each conduction and valence band is two-fold degenerate. Due to the anti-commutativity of the above Γ matrices, we also obtain a simple expression for the bare fermion propagator [cf. Eq. (5.58)]:

$$\mathcal{G}_0(k) = (-i\nu \mathbb{1} + \mathbf{d}(\mathbf{k}) \cdot \boldsymbol{\Gamma})^{-1} = \frac{i\nu \mathbb{1} + \mathbf{d}(\mathbf{k}) \cdot \boldsymbol{\Gamma}}{\nu^2 + |\mathbf{d}(\mathbf{k})|^2} \quad (5.65)$$

Furthermore, as the Γ matrices are traceless, *i.e.*, $\text{tr} \Gamma_a = 0$ for all $i = 1, \dots, 5$, we can easily calculate the matrix trace in Eq. (5.59). As an intermediate result, we obtain:

$$\Pi_{\text{FKM}}(q) = 8 \int \frac{d^4 k}{(2\pi)^4} \frac{-\nu(\nu + \omega) + \mathbf{d}(\mathbf{k}) \cdot \mathbf{d}(\mathbf{k} + \mathbf{q})}{[\nu^2 + |\mathbf{d}(\mathbf{k})|^2][(\nu + \omega)^2 + |\mathbf{d}(\mathbf{k} + \mathbf{q})|^2]}. \quad (5.66)$$

To compute the frequency integral over $-\infty < \nu < \infty$, we perform a rotation to the imaginary frequency axis by substituting $\nu = iz$. In the static limit, *i.e.*, for $\omega = 0$, we then find the following expression for the polarization function using the well-known residue theorem:

$$\Pi_{\text{FKM}}(\omega = 0, \mathbf{q}) = 4 \int \frac{d^3 k}{(2\pi)^3} \frac{|\mathbf{d}(\mathbf{k})| |\mathbf{d}(\mathbf{k} + \mathbf{q})| - \mathbf{d}(\mathbf{k}) \cdot \mathbf{d}(\mathbf{k} + \mathbf{q})}{|\mathbf{d}(\mathbf{k})| |\mathbf{d}(\mathbf{k} + \mathbf{q})| (|\mathbf{d}(\mathbf{k})| + |\mathbf{d}(\mathbf{k} + \mathbf{q})|)}. \quad (5.67)$$

Concerning the role of the spin-orbit coupling λ_{SO} we observe that the bulk band gap remains finite as long as we keep λ_{SO}/t finite and include modulations of the hopping amplitude, δt_j [cf. Figs. 5.12, 5.13, and 5.17]. In that case, we may perform a Taylor expansion of the integrand for low-energy excitations with small wave vector \mathbf{q} :

$$\mathbf{d}(\mathbf{k} + \mathbf{q}) \approx \mathbf{d}(\mathbf{k}) + q_\alpha \partial_\alpha \mathbf{d}(\mathbf{k}) + \frac{1}{2!} q_\alpha q_\beta \partial_\alpha \partial_\beta \mathbf{d}(\mathbf{k}) + \mathcal{O}(q^3), \quad (5.68a)$$

and the Taylor expansion of the vector norm with respect to \mathbf{q} is given by:

$$\begin{aligned}
|\mathbf{d}(\mathbf{k} + \mathbf{q})| &\approx |\mathbf{d}(\mathbf{k})| + q_\alpha \partial_\alpha |\mathbf{d}(\mathbf{k})| + \frac{1}{2!} q_\alpha q_\beta \partial_\alpha \partial_\beta |\mathbf{d}(\mathbf{k})| + \mathcal{O}(q^3) \\
&= |\mathbf{d}(\mathbf{k})| + q_\alpha \frac{d_\beta(\mathbf{k}) \partial_\alpha d_\beta(\mathbf{k})}{|\mathbf{d}(\mathbf{k})|} \\
&\quad + \frac{q_\alpha q_\beta}{2} \frac{(\partial_\alpha d_\gamma(\mathbf{k})) (\partial_\beta d_\gamma(\mathbf{k})) + d_\gamma(\mathbf{k}) \partial_\alpha \partial_\beta d_\gamma(\mathbf{k})}{|\mathbf{d}(\mathbf{k})|} \\
&\quad - \frac{q_\alpha q_\beta}{2} \frac{d_\gamma(\mathbf{k}) (\partial_\alpha d_\gamma(\mathbf{k})) d_\delta(\mathbf{k}) (\partial_\alpha d_\delta(\mathbf{k}))}{|\mathbf{d}(\mathbf{k})|^3} + \mathcal{O}(q^3),
\end{aligned} \tag{5.68b}$$

where $\partial_\alpha f(\mathbf{k}) \equiv \partial f(\mathbf{k}) / \partial k_\alpha$, and a summation over the Greek indices $\alpha, \beta, \gamma, \delta \in \{x, y, z\}$ is implied. Upon substituting the above expressions, the static polarization function takes the following form:

$$\begin{aligned}
\Pi_{\text{FKM}}(\omega = 0, \mathbf{q}) &\approx 2q_\alpha q_\beta \int \frac{d^3 k}{(2\pi)^3} \left[\frac{(\partial_\alpha d_\gamma(\mathbf{k})) (\partial_\beta d_\gamma(\mathbf{k}))}{|\mathbf{d}(\mathbf{k})|^3} \right. \\
&\quad \left. - \frac{d_\gamma(\mathbf{k}) (\partial_\alpha d_\gamma(\mathbf{k})) d_\delta(\mathbf{k}) (\partial_\beta d_\delta(\mathbf{k}))}{|\mathbf{d}(\mathbf{k})|^5} \right] + \mathcal{O}(q^3), \tag{5.69}
\end{aligned}$$

or in a more compact notation:

$$\Pi_{\text{FKM}}(\omega = 0, \mathbf{q}) \approx 2 \int \frac{d^3 k}{(2\pi)^3} \left\{ \frac{[(\mathbf{q} \cdot \nabla_{\mathbf{k}}) \mathbf{d}(\mathbf{k})]^2}{|\mathbf{d}(\mathbf{k})|^3} - \frac{[\mathbf{d}(\mathbf{k}) \cdot (\mathbf{q} \cdot \nabla_{\mathbf{k}}) \mathbf{d}(\mathbf{k})]^2}{|\mathbf{d}(\mathbf{k})|^5} \right\} + \mathcal{O}(q^3). \tag{5.70}$$

This result for the static polarization function in the limit of long wave lengths, $\mathbf{q} \rightarrow 0$, is apparently well-defined as long as the bulk remains gapped which is ensured by a finite spin-orbit coupling $\lambda_{\text{SO}}/t \neq 0$ and finite hopping modulations δt_j , as mentioned before. Furthermore, from our previous analysis of the bulk band structure and density of states we also know that in the limit $\lambda_{\text{SO}}/t = 0$ nodal lines emerge in the bulk of system close to the X points, and those nodal lines also give rise to the perfectly flat surface states. Close to the bulk quantum critical point $\lambda_{\text{SO}}/t = 0$, we can describe such a nodal line by an effective low-energy Hamiltonian

$$\mathcal{H}_{\text{eff}}(\mathbf{k}) = \tilde{\mathbf{d}}(\mathbf{k}) \cdot \boldsymbol{\tau} \quad \text{with} \quad \tilde{\mathbf{d}}(\mathbf{k}) = (k_x, k_y, \lambda_{\text{SO}})^T, \tag{5.71}$$

where we have performed a suitable coordinate transformation, so that k_z is the momentum along the nodal line which does not enter the effective Hamiltonian. Upon substituting the vector $\tilde{\mathbf{d}}(\mathbf{k})$ into the polarization function $\Pi(\omega = 0, \mathbf{q})$ [cf. Eq. (5.70)] we find that

$$\Pi_{\text{eff}}(\omega = 0, \mathbf{q}) \propto \frac{q_x^2 + q_y^2}{|\lambda_{\text{SO}}|}. \tag{5.72}$$

In the Fu-Kane-Mele model, there are three equivalent X points in the Brillouin zone of the diamond lattice, located $X_x = \frac{2\pi}{a} (1, 0, 0)^T$, $X_y = \frac{2\pi}{a} (0, 1, 0)^T$, and $X_z = \frac{2\pi}{a} (0, 0, 1)^T$. Hence, combining the above result for all three X points we find that the static polarization function develops a $1/\lambda_{\text{SO}}$ divergence close to the bulk critical point:

$$\Pi_{\text{FKM}}(\omega = 0, \mathbf{q}) \propto \frac{|\mathbf{q}|^2}{|\lambda_{\text{SO}}|} + \mathcal{O}(q^3). \tag{5.73}$$

Note that the above result was obtained in the absence of next-nearest neighbor hopping t' . This particular form of the polarization function is reminiscent of the situation in *two*-dimensional graphene sheets with a single Dirac *point* and mass m , where the static polarization function takes the following form^{122,155}:

$$\Pi_{\text{graphene}}(\omega = 0, \mathbf{q}) \propto \frac{|\mathbf{q}|^2}{\sqrt{v_F^2 |\mathbf{q}|^2 + m^2}}. \quad (5.74)$$

According to our discussion, we find a similar relation in the *three*-dimensional Fu-Kane-Mele model for a *line* of Dirac points, where the spin-orbit interaction generates a mass term for the otherwise gapless Dirac fermions:

$$\Pi_{\text{FKM}}(\omega = 0, \mathbf{q}) \propto \frac{|\mathbf{q}|^2}{\sqrt{v_F^2 |\mathbf{q}|^2 + \lambda_{\text{SO}}^2}} \propto \begin{cases} |\mathbf{q}|^2/|\lambda_{\text{SO}}| & \text{for } \lambda_{\text{SO}} \neq 0 \\ |\mathbf{q}|/v_F & \text{for } \lambda_{\text{SO}} = 0 \end{cases} \quad (5.75)$$

An important consequence of this result is that the dielectric constant ϵ [cf. Eq. (5.60)] diverges upon approaching the bulk quantum critical point $\lambda_{\text{SO}}/t = 0$, which can be interpreted a precursor to the screening of long-range Coulomb interactions by the bulk nodal lines:

$$\epsilon = 1 + \frac{4\pi e}{|\mathbf{q}|^2} \Pi(\omega = 0, \mathbf{q}) \propto c_0 + \frac{c_1}{|\lambda_{\text{SO}}|} \rightarrow \infty \quad \text{as } \lambda_{\text{SO}} \rightarrow 0 \quad (5.76)$$

with some numerical constants $c_{0,1}$. This particular divergence of the dielectric constant has a profound consequence on the effective interaction strength α :

$$\alpha = \frac{e^2}{\hbar v_F^{\text{surf}} \epsilon} \propto \frac{1}{|\lambda_{\text{SO}}|} \frac{1}{c_0 + c_1/|\lambda_{\text{SO}}|} = \frac{1}{c_0 |\lambda_{\text{SO}}| + c_1} \rightarrow \frac{1}{c_1} \quad \text{for } \lambda_{\text{SO}} \rightarrow 0. \quad (5.77)$$

Here, we have used that the surface Fermi velocity close to the bulk critical point is roughly given by $v_F^{\text{surf}} \propto |\lambda_{\text{SO}}|$ if next-nearest neighbor hopping is neglected (cf. discussion of the surface bands in section 5.3.3). We expect that this result is not changed qualitatively when hopping between next-nearest neighbors is included. Most importantly, note that the effective interaction strength does not generically become large when $v_F^{\text{surf}} \rightarrow 0$ upon approaching the quantum critical point, in stark contrast to our naive expectation that a flat surface band leads to a diverging interaction strength α , thereby opening a surface band gap (see section 5.1). Hence, in general we do not expect a spontaneous mass generation due to long-range Coulomb interactions in the Fu-Kane-Mele model and related models either due to the screening of Coulomb interactions.

5.3.6 Discussion

To conclude, we have shown that the existence and topological stability of the flat surface bands in the Fu-Kane-Mele model are guaranteed by the nodal lines in the bulk of the system present at $\lambda_{\text{SO}}/t = 0$, whose projections onto the 2D surfaces of a finite system determine the range of the flat surface states. However, in general we expect that those nodal lines also lead to screening of the long-range Coulomb interactions as the static polarization function $\Pi(\omega = 0, \mathbf{q}) \propto |\mathbf{q}|^2/|\lambda_{\text{SO}}|$ shows a particular divergence as a function of the spin-orbit coupling λ_{SO} . Concerning the effective interaction strength $\alpha \propto 1/(v_F^{\text{surf}} \epsilon)$, we find that the divergence of the dielectric constant $\epsilon \propto c_0 + c_1/|\lambda_{\text{SO}}|$ is compensated by a reduction of the Fermi

velocity of the surface Dirac fermions as $v_F^{\text{surf}} \propto \lambda_{\text{SO}}$. If we take the hopping t' between second-nearest neighbors into account, the chiral symmetry of the Fu-Kane-Mele model is broken, and the surface states become dispersive with a reduced surface Fermi velocity $v_F^{\text{surf}} = (t'/t)v_F^{\text{bulk}}$, similar to the situation in the Kane-Mele model on the graphene lattice (see previous section). In any case, we expect to find surface states with small Fermi velocity, but at the same time the dielectric constant is expected to diverge linearly as function of the spin-orbit coupling λ_{SO} , which is a precursor to screening. Consequently, the effective interaction strength α does not become large close to the critical point, but is expected to be of the same order as other energy scales, and chiral symmetry breaking due to long-range Coulomb interactions is not expected in general. Our analysis also suggests that probably the most promising situation to observe a spontaneously generated surface mass in the Fu-Kane-Mele model could be a situation, where the bulk band gap is still pronounced and screening is not yet too strong, but here it comes down to numbers. Furthermore, as the average kinetic energy of the surface Dirac fermions goes to zero as we approach the bulk quantum critical point, even local interactions of arbitrarily small strength are expected to create a gap, as we discuss in section 5.5 on a mean-field level.

5.4 Flat surface bands in the minimal model for HgTe

In the previous section, we have discussed the properties of the Fu-Kane-Mele model, and we have seen that chiral symmetry breaking due to long-range Coulomb interactions will not generically occur in that model, although the corresponding surface states become flat upon approaching the bulk quantum critical point. However, this model is somewhat special in the sense that it describes the spin-orbit interactions in terms of a next-nearest neighbor hopping due to the s orbital nature of the electronic states instead of an intrinsic or atomic spin-orbit coupling of p orbital states. Moreover, the diamond lattice has a large number of crystal symmetries protecting the topological properties of the sample. For example, the Fu-Kane-Mele model is a semimetal with zero band gap in the presence of uniform overlap parameters, and we have to increase the overlap parameter along one of the bonds, so that the Fu-Kane-Mele model becomes a strong topological insulator, as discussed above. Also, the trivial band insulator phase is not perturbatively accessible in the Fu-Kane-Mele model unless we take a staggered sublattice potential into account or drastically increase the hopping amplitude along one of the bonds⁷.

Another class of topological insulator models is exemplified by the minimal model for topological insulators in three dimensions which we have introduced in chapter 3 (Ref. 19). In contrast to the Fu-Kane-Mele model, this effective model is based on superpositions of the p orbital states and thus allows for intrinsic spin-orbit interactions without the need of next-nearest neighbor interactions. Furthermore, the corresponding phase diagram contains both strong and weak topological phases as well as a trivial band insulator phase. However, it turns out that in the absence of inversion symmetry this model also exhibits an intermediate metallic phase in which the bulk Dirac points shift around in the bulk Brillouin zone as function of the tuning parameter until they meet a time-reversed partner and annihilate. Such a phase is also known as a **Weyl metal**, because the effective low-energy theory can be described in terms of gapless Weyl fermions^{58,180–182}. We also find that the polarization function develops a divergence as a precursor to screening of long-ranged Coulomb interactions due to a line of nodal points in the bulk of system, as we discuss below. Consequently, also in this model the effective interaction strength α does not diverge, and we do not find chiral symmetry breaking in that model.

In the following, we investigate the behavior of the surface bands in the minimal model without inversion symmetry by solving the tight-binding Hamiltonian for a slab geometry with (001) oriented faces. We show that the Fermi velocity v_F^{surf} of the surface Dirac fermions also vanishes upon approaching the bulk quantum critical point, because the bulk conduction and valence bands form a “spectral stamp” squeezing the surface bands. As a consequence of this “spectral pressure” the Fermi velocity v_F^{surf} of the surface Dirac fermions becomes smaller. After that, we discuss the polarization function in the zero-frequency and long-wavelength limit and compute the dielectric constant ϵ as for the Fu-Kane-Mele model. In particular, we show that close to the quantum critical point ϵ diverges due to the appearance of the bulk nodal lines. Similar to the Fu-Kane-Mele model, this divergence renders the effective interaction strength α small enough, so that a spontaneous mass generation is in general avoided in the minimal model for 3D topological insulators as well.

5.4.1 Flat surface bands in the minimal model

In chapter 3, we have introduced an effective tight-binding model for 3D topological insulators which was based on the bulk bands close to the Fermi level in strained 3D HgTe. After performing a Fourier decomposition of this model on the simple cubic lattice, the corresponding Bloch Hamiltonian was formulated as follows:

$$\mathcal{H}(\mathbf{k}) = m \Gamma_0 - t \sum_{j=1}^3 (\cos(k_j a) \Gamma_0 + \sin(k_j a) \Gamma_j) + \Delta' \Gamma_{04}. \quad (5.78)$$

Here, m is the **mass parameter** or **tuning parameter** of the minimal model which allows to realize different quantum states of matter, notably a trivial band insulator phase for $|m| > 3t$ and a strong topological insulator phase for $t < |m| < 3t$ (see Fig. 3.17). Furthermore, we consider the 4×4 matrices Γ_a and their commutators Γ_{ab} as introduced in chapter 3. Note that we explicitly take the term $\Delta' \Gamma_{04}$ into account, which breaks inversion symmetry, but does not break time-reversal symmetry. Diagonalizing the above Bloch Hamiltonian we find the bulk electronic band structure:

$$\mathcal{E}_{\pm\pm}(\mathbf{k}) = \pm \sqrt{d_0(\mathbf{k})^2 + \left(\sqrt{d_1(\mathbf{k})^2 + d_2(\mathbf{k})^2 + d_3(\mathbf{k})^2} \pm d_{04}(\mathbf{k}) \right)^2} \quad (5.79)$$

where the tight-binding functions $d_j(\mathbf{k})$ are defined as:

$$d_0(\mathbf{k}) \equiv m - t(\cos(k_x a) + \cos(k_y a) + \cos(k_z a)), \quad (5.80a)$$

$$d_1(\mathbf{k}) \equiv -t \sin(k_x a), \quad d_2(\mathbf{k}) \equiv -t \sin(k_y a), \quad d_3(\mathbf{k}) \equiv -t \sin(k_z a), \quad (5.80b)$$

and

$$d_{04}(\mathbf{k}) \equiv \Delta' = \text{const.} \quad (5.80c)$$

As shown in chapter 3, in an inversion-symmetric system, *i.e.*, for $\Delta'/t = 0$, the valence and conduction bands are doubly degenerate, and the bulk band gap closes at one of the **time-reversal invariant momenta** Γ , X , M , or R in the Brillouin zone, depending on the value of the tuning parameter m/t .

However, the situation is much more interesting in system with broken inversion symmetry, where the two-fold degeneracy between the valence and conduction bands is lifted due to a finite term $\Delta' \Gamma_{04} \neq 0$. To close the bulk band in such systems, one needs to fulfill two conditions

simultaneously:

$$m/t \doteq \cos(k_x a) + \cos(k_y a) + \cos(k_z a), \quad (5.81a)$$

$$\Delta'/t \doteq \sqrt{\sin^2(k_x a) + \sin^2(k_y a) + \sin^2(k_z a)}. \quad (5.81b)$$

Each of the two conditions defines a two-dimensional manifold of solutions in momentum space, and for given values of the tuning parameter m/t and inversion asymmetry Δ'/t we find three possible situations:

- (i) If there is no point in momentum space which fulfills both conditions at a time, then the system describes either a trivial insulator or a topological insulator. Those phases can be easily distinguished by their different \mathbb{Z}_2 invariants $(\nu_0; \nu_1\nu_2\nu_3)$.
- (ii) If, however, we find isolated points fulfilling both relations, then we are either at the bulk quantum critical point or in the intermediate metallic Weyl phase. Note that both the critical point and the metallic Weyl phase are characterized by isolated bulk points with linear dispersion.
- (iii) If we find lines of solutions in momentum space, then we are also in the intermediate metallic Weyl phase. Those lines are, however, an artifact of the system being particle-hole symmetric, and in a generic system the lines will be reduced to single bulk Dirac points.

As a result, for an inversion-asymmetric system with $\Delta'/t \neq 0$ we may identify three different phases: $m < m_{c1}$ and $m > m_{c2}$ describe the topological insulator and trivial band insulator phases, respectively. For $m_{c1} < m < m_{c2}$, however, the system enters an intermediate metallic phase, whose size depends on the inversion asymmetry Δ'/t . The value of m_{c1} can be estimated from a gradient expansion of the tight-binding functions $d_j(\mathbf{k})$ in the limit $\Delta'/t \ll 1$ which leads to

$$\frac{m_{c1}}{t} = 3 - \frac{1}{2} \left(\frac{\Delta'}{t} \right)^2. \quad (5.82)$$

It is important to note that at $m = m_{c1}$ the bulk band gap closes, but this happens away from the time-reversal invariant momentum $\Gamma = (0, 0, 0)^T$ at a distance set by $|\mathbf{k}| = \Delta'/t$. Fig. 5.23 shows the closing of the bulk band gap within the linearized band structure of the minimal model close to the critical point $m \lesssim m_{c1}$. Although the bulk band gap appears to close on a sphere $|\mathbf{k}| = \Delta'/t$, this is an artifact of the low-order expansion, and higher-order terms induce cubic anisotropies which reduce the spherical symmetry to isolated Dirac points. Nevertheless, the distance of the bulk Dirac points is set by Δ'/t .

Now let us consider the surface states of such an inversion-asymmetric system. Approaching the bulk quantum critical point m_{c1} from within the strong topological insulator phase, *i.e.*, for $m < m_{c1}$, the bulk Dirac cones are gapped, and the bulk band gap is given by $\mathcal{E}_{\text{gap}} = 2 |\delta m| = 2 |m - m_{c1}|$. For an arbitrary surface with normal vector \hat{n} we can project the bulk band gap onto the corresponding surface Brillouin zone. Fig. 5.24 shows the projected bulk conduction and valence bands for a (001) surface with normal vector in the [001] direction, and at the critical point m_{c1} the bulk band gap closes. Consequently, in a finite-size sample, the surface bands connect states originating from the bulk across the gap, and due to Kramers' theorem those surface bands have to cross at the center of the surface Brillouin zone, $\bar{\Gamma} = (0, 0)^T$, as projection of the time-reversal invariant momentum Γ . In that sense, we find that in the vicinity of the critical point $m = m_{c1}$ the bulk bands exert a ‘‘spectral pressure’’ onto the surface bands, and thus the surface bands become more and more flat as $m \rightarrow m_{c1}$. A rough estimate

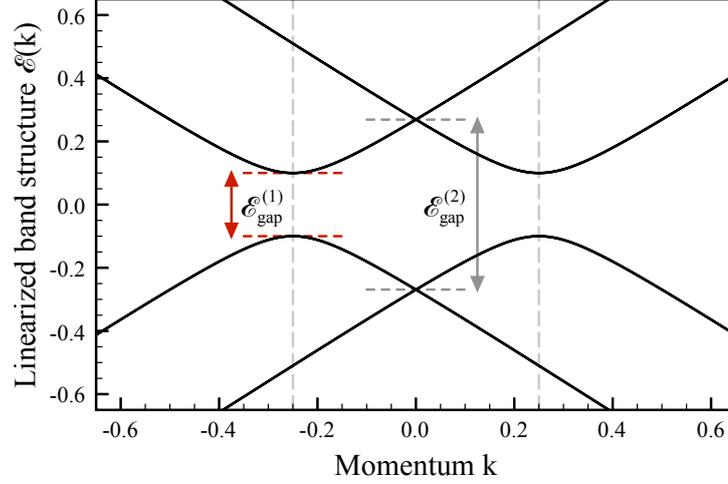


Figure 5.23. Plot of the linearized band structure $\mathcal{E}_{\pm\pm}(\mathbf{k})$ of the minimal model realizing a strong topological insulator for $t = 1$, $m/t = 2.9$, and $\Delta'/t = 0.25$. The system is characterized by two gap scales $\mathcal{E}_{\text{gap}}^{(1)} = 2|\delta m|$ and $\mathcal{E}_{\text{gap}}^{(2)} = 2\sqrt{\delta m^2 + \Delta'^2}$, where $\delta m \equiv m - m_{c1}$ is the distance to the quantum critical point m_{c1} . Upon approaching the bulk quantum critical point the gap $\mathcal{E}_{\text{gap}}^{(1)}$ closes at $k = \pm\Delta'/t$, while $\mathcal{E}_{\text{gap}}^{(2)}$ remains finite. As a consequence, close to the bulk quantum critical point the two bulk bands $\mathcal{E}_{\pm,-}(\mathbf{k})$ lead to singular contributions to the static polarization function $\Pi(\mathbf{q}, \omega = 0)$, as discussed in the main text.

for the Fermi velocity v_F^{surf} of the surface Dirac fermions can be obtained from a simple linear interpolation between the bulk Dirac cones at the edges of the projected bulk band gap:

$$v_F^{\text{surf}} \approx \frac{\mathcal{E}_{\text{gap}}}{\Delta'/t} \propto v_F^{\text{bulk}} \frac{|\delta m|}{\Delta'}, \quad (5.83)$$

where $v_F^{\text{bulk}} = ta$ denotes the Fermi velocity of the bulk fermions. Obviously, the Fermi velocity v_F^{surf} can be reduced drastically when approaching the bulk quantum critical point $m = m_{c1}$, but the surface bands do not become perfectly flat in that limit. This can be understood as follows: Although the projected bulk band gap exerts a spectral pressure onto the surface bands, higher-order terms invalidate the picture of a perfectly flat stamp, but introduce a roughening of the stamp surface on the scale $(\Delta'/t)^2$ (see Fig. 5.24). Consequently, the bulk band gap closes only at isolated bulk Dirac points, and the surface bands can bend up- or downwards, so that they do not become perfectly flat anymore.

To confirm this intuitive picture, we have calculated the surface states of the minimal model by solving the Bloch Hamiltonian in a finite slab geometry with open boundary conditions in the z direction of the (001) slab and periodic boundary conditions in x and y directions:

$$\begin{aligned} \mathcal{H}(\mathbf{k}_{\parallel}) = & \sum_z \Psi_{\mathbf{k}_{\parallel}z}^{\dagger} (m \Gamma_0 + \Delta' \Gamma_{04}) \Psi_{\mathbf{k}_{\parallel}z} \\ & - t \sum_z \Psi_{\mathbf{k}_{\parallel}z}^{\dagger} (\cos(k_x a) + \cos(k_y a)) \Gamma_0 \Psi_{\mathbf{k}_{\parallel}z} \\ & - t \sum_z \Psi_{\mathbf{k}_{\parallel}z}^{\dagger} (\sin(k_x a) \Gamma_1 + \sin(k_y a) \Gamma_2) \Psi_{\mathbf{k}_{\parallel}z} \\ & - t \sum_z \left[\Psi_{\mathbf{k}_{\parallel}z}^{\dagger} \left(\frac{\Gamma_0 - i\Gamma_3}{2} \right) \Psi_{\mathbf{k}_{\parallel},z-a}^{\dagger} + \text{H.c.} \right]. \end{aligned} \quad (5.84)$$

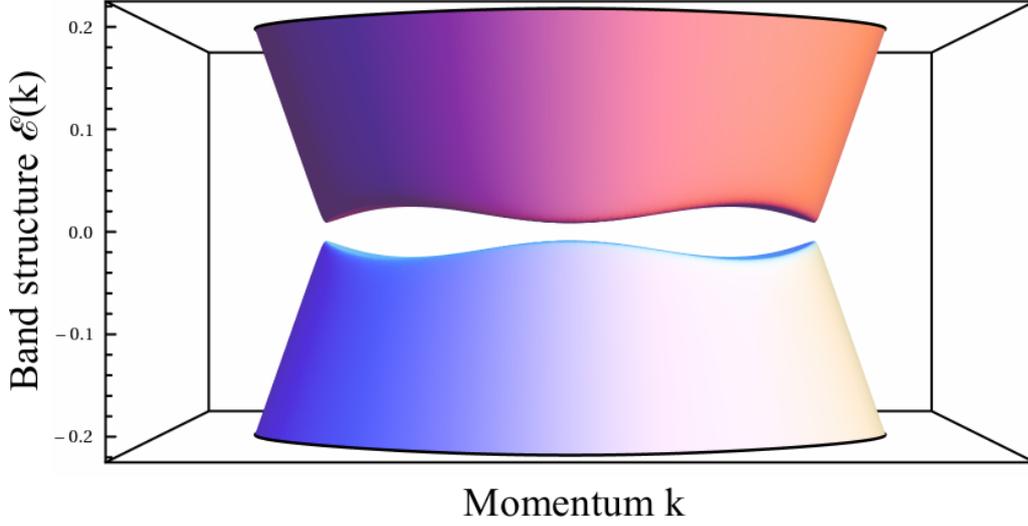


Figure 5.24. Projection of the bulk conduction and valence bands onto the surface Brillouin zone of a (001) surface. To leading order, the bulk bands have the form of a stamp which closes as $m \rightarrow m_{c1}$. Higher-order corrections lead to a fine structure of the projected bulk bands which is clearly visible in the Figure.

Fig. 5.25 shows the two-dimensional slab band structure for a slab with (001) surfaces along the high-symmetry lines of the surface Brillouin zone, while Fig. 5.26 shows a magnification of the slab band structure close to the center of the Brillouin zone. Apparently, the nearby bulk bands exert pressure on the surface bands, so that they become increasingly flat upon approaching the bulk quantum critical point which confirms our basic understanding of the flatness of the surface bands in the minimal model.

To conclude this section, we have established a possible route to control the Fermi velocity of the surface Dirac fermions appearing in the minimal model for 3D topological insulators by tuning m as $v_F^{\text{surf}} \propto |m - m_{c1}|/t$. However, due to the nearby presence of the bulk bands we also have to consider screening mechanisms in that model. Although the projected bulk gap in the form of a stamp exerts spectral pressure onto the surface bands and forces them to become quite flat, the same bulk stamp can also lead a finite density of states on an intermediate energy scale which eventually may lead to screening of the long-range Coulomb interaction.

5.4.2 Polarization function for the minimal model

Let us briefly discuss the polarization function for the minimal model of a 3D topological insulator close to the bulk quantum critical point $m/t = 3$ in the presence of inversion asymmetry. For $\delta m/t \equiv |m - 3t|/t \ll 1$ we may consider a linearized version of the minimal model (5.78),

$$\mathcal{H}_{\text{eff}}(\mathbf{k}) \approx \delta m \Gamma_0 - at \sum_{j=1}^3 k_j \Gamma_j + \Delta' \Gamma_{04}, \quad (5.85)$$

where \mathbf{k} measures the distance to the center of the Brillouin zone, *i.e.*, at the Γ point. As introduced before, we define the free Green's function as the single-particle propagator

$$\mathcal{G}_0(k) \equiv \langle \Psi_k^\dagger \Psi_k \rangle_0 = (-i\nu \mathbb{1} + \mathcal{H}_{\text{eff}}(\mathbf{k}))^{-1}, \quad (5.86)$$

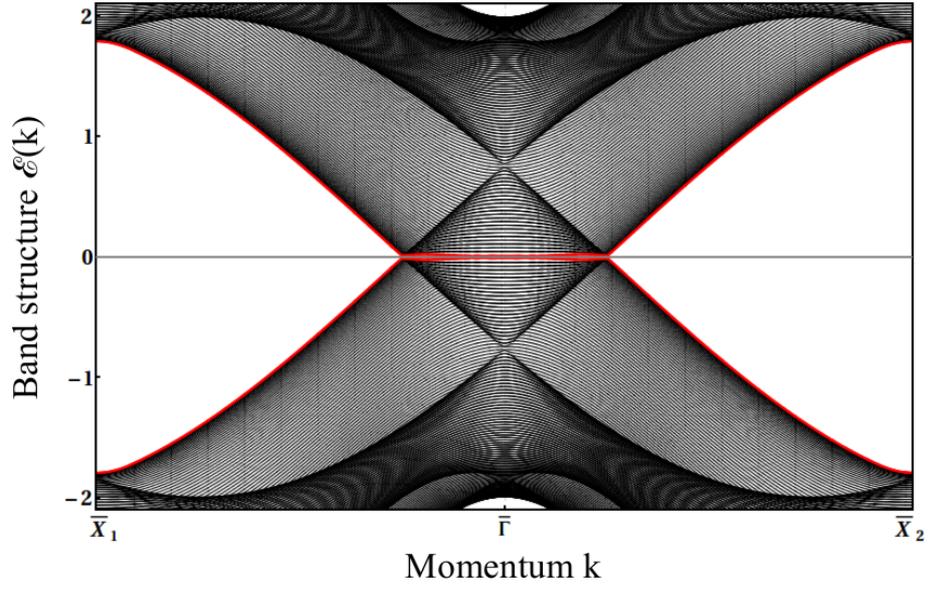


Figure 5.25. Plot of the electronic slab band structure for a (001) slab consisting of $N_z = 100$ layers. The parameters of the minimal model (5.84) are chosen such that it describes an inversion-asymmetric topological insulator: $t = 1$ and $m/t = 2.7$. The inversion asymmetry $\Delta'/t = 0.7$ was chosen artificially large to show the appearance of the bulk band gap exerting spectral pressure onto the surface bands indicated by red lines.

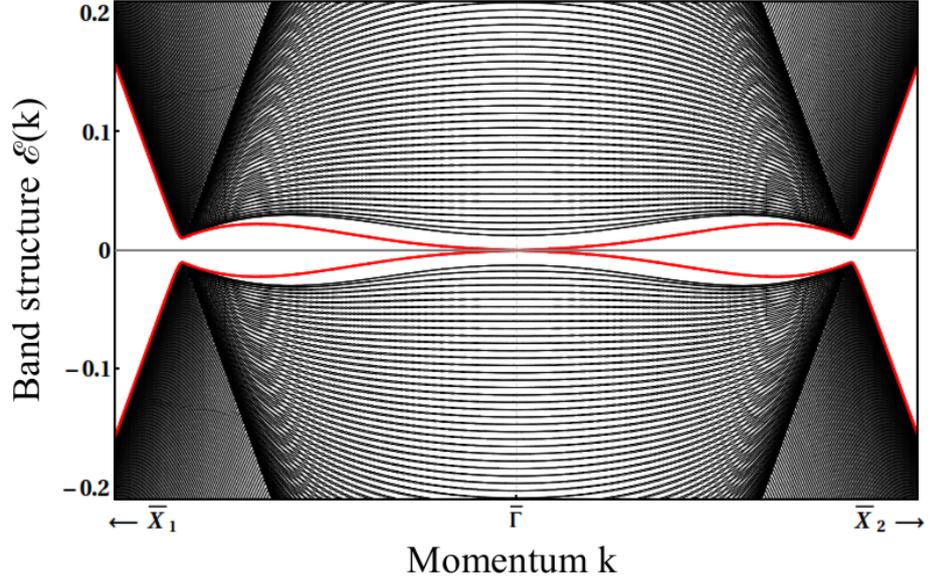


Figure 5.26. Zoom into the electronic slab band structure for a (001) slab consisting of $N_z = 500$ layers. The parameters of the minimal model are identical to those in Fig. 5.25. The inversion asymmetry Δ' shifts the bulk Dirac points away from the time-reversal invariant momentum Γ on a scale set by $|\mathbf{k}| = \Delta'/t$ and leads to the formation of a stamp-shaped bulk band gap. The surface of the stamp is roughened on a scale $(\Delta'/t)^2$ which allows the surface bands (indicated by red lines) to bend up- or downwards and to interact with the bulk states by elastic scattering. A simple estimate for the Fermi velocity of the surface Dirac fermions can be given as $v_F^{\text{surf}} \approx \mathcal{E}_{\text{gap}}/(\Delta'/t) \propto v_F^{\text{bulk}}(|m - m_{c1}|/\Delta')$, where $v_F^{\text{bulk}} = at$ denotes the bulk Fermi velocity.

where $k \equiv (\nu, \mathbf{k})$ is the 4-momentum consisting of both frequency ν and momentum \mathbf{k} . Let $\mathcal{U}(\mathbf{k})$ be the unitary matrix of normalized eigenvectors of the effective Hamiltonian $\mathcal{H}_{\text{eff}}(\mathbf{k})$, i.e.,

$$\mathcal{U}(\mathbf{k})\mathcal{H}_{\text{eff}}(\mathbf{k})\mathcal{U}(\mathbf{k})^\dagger = \text{diag}(\{\mathcal{E}_n(\mathbf{k})\}). \quad (5.87)$$

Here, $\mathcal{E}_n(\mathbf{k})$ denotes the eigenenergies of $\mathcal{H}_{\text{eff}}(\mathbf{k})$ in ascending order:

$$\mathcal{E}_{\pm\pm}(\mathbf{k}) \equiv \pm\sqrt{\delta m^2 + (at|\mathbf{k}| \pm \Delta')^2}. \quad (5.88)$$

As a consequence, the free Green's function $\mathcal{G}_0(k)$ can be rewritten as

$$\mathcal{G}_0(k) = \mathcal{U}(\mathbf{k})\mathcal{D}(k)\mathcal{U}(\mathbf{k})^\dagger \quad \text{with} \quad \mathcal{D}(k) \equiv \text{diag}(\{(-i\nu + \mathcal{E}_n(\mathbf{k}))^{-1}\}). \quad (5.89)$$

In analogy to the discussion of the Fu-Kane-Mele model in the previous section, we consider the polarization function $\Pi(q)$ as defined in Eq. (5.59). Upon substituting the free Green's function $\mathcal{G}_0(k)$ we can rewrite the above relation for the polarization function as

$$\Pi(q) = 2 \int \frac{d^4k}{(2\pi)^4} \text{tr}\{\mathcal{F}(\mathbf{k}, \mathbf{q})^\dagger \mathcal{D}(k) \mathcal{F}(\mathbf{k}, \mathbf{q}) \mathcal{D}(k + \mathbf{q})\}, \quad (5.90)$$

where, for brevity, we have introduced a **structure factor** $\mathcal{F}(\mathbf{k}, \mathbf{q})$ which depends on the momenta \mathbf{k} and \mathbf{q} as follows:

$$\mathcal{F}(\mathbf{k}, \mathbf{q}) \equiv \mathcal{U}(\mathbf{k})^\dagger \mathcal{U}(\mathbf{k} + \mathbf{q}). \quad (5.91)$$

Since $\mathcal{D}(k)$ is diagonal, the polarization function $\Pi(q)$ then takes the form

$$\Pi(q) = 2 \int \frac{d^4k}{(2\pi)^4} \sum_{m,n=1}^4 \frac{[\mathcal{F}(\mathbf{k}, \mathbf{q})^\dagger]_{mn} \mathcal{F}(\mathbf{k}, \mathbf{q})_{nm}}{(ik_0 - \mathcal{E}_n(\mathbf{k}))(ik_0 + iq_0 - \mathcal{E}_m(\mathbf{k} + \mathbf{q}))}. \quad (5.92)$$

First, let us perform the integration over the internal frequency, ν . As a result, we obtain:

$$\Pi(q) = 2 \int \frac{d^3k}{(2\pi)^3} \sum_{m,n=1}^4 [\mathcal{F}(\mathbf{k}, \mathbf{q})^\dagger]_{mn} \mathcal{F}(\mathbf{k}, \mathbf{q})_{nm} \frac{n_F(\mathcal{E}_n(\mathbf{k})) - n_F(\mathcal{E}_m(\mathbf{k} + \mathbf{q}) - iq_0)}{(iq_0 - (\mathcal{E}_m(\mathbf{k} + \mathbf{q}) - \mathcal{E}_n(\mathbf{k}))). \quad (5.93)$$

At zero temperature, all states below (above) the chemical potential μ are occupied (empty) since $n_F(\epsilon - \mu) = \Theta(\mu - \epsilon)$ with $\Theta(x)$ denoting the Heaviside theta function, so that the static polarization function takes the following form ($\mu = 0$):

$$\Pi(\omega = 0, \mathbf{q}) = 2 \int \frac{d^3k}{(2\pi)^3} \sum_{m,n=1}^4 [\mathcal{F}(\mathbf{k}, \mathbf{q})^\dagger]_{mn} \mathcal{F}(\mathbf{k}, \mathbf{q})_{nm} \frac{\Theta(-\mathcal{E}_m(\mathbf{k} + \mathbf{q})) - \Theta(-\mathcal{E}_n(\mathbf{k}))}{\mathcal{E}_m(\mathbf{k} + \mathbf{q}) - \mathcal{E}_n(\mathbf{k})}. \quad (5.94)$$

Here, the only non-vanishing contributions to the static polarization function stem from particle-hole excitations:

$$\Theta(-\mathcal{E}_m(\mathbf{k} + \mathbf{q})) - \Theta(-\mathcal{E}_n(\mathbf{k})) = \begin{cases} 1 & m \text{ occupied and } n \text{ empty} \\ -1 & m \text{ empty and } n \text{ occupied} \\ 0 & \text{both } m \text{ and } n \text{ empty or occupied} \end{cases} \quad (5.95)$$

Note that the energy denominator $\mathcal{E}_m(\mathbf{k} + \mathbf{q}) - \mathcal{E}_n(\mathbf{k})$ is non-zero as long as $\delta m \neq 0$, because the indices m and n refer to either valence or conduction bands, respectively.

Concerning the quantum critical point of the minimal model we find that the static polarization function (5.93) develops a singularity in the long-wavelength limit at the bulk quantum critical point. In the following, we consider the inversion asymmetry $\Delta' \Gamma_{04}$ as a small perturbation to the strong topological insulator, *i.e.*, we focus on the case $\Delta'/t \ll 1$, and focus on the leading-order terms of $\Pi(\omega = 0, \mathbf{q})$ in the limit $\mathbf{q} \rightarrow 0$. Starting from the linearized band structure (5.88) we see that in the vicinity of the bulk Dirac point the singular contribution to the polarization function stems from the touching of the highest valence band and the lowest conduction band:

$$\mathcal{E}_{\pm,-}(\mathbf{k}) = \pm \sqrt{\delta m^2 + (at|\mathbf{k}| - \Delta')^2}, \quad (5.96)$$

where, for simplicity, we have assumed that Δ' is positive. The divergence of the polarization is due to the fact that in the limit $\mathbf{q} \rightarrow 0$ and $\delta m \rightarrow 0$ the energy denominator entering Eq. (5.93) vanishes. Note that the other terms entering Eq. (5.93) remain constant or vanish quadratically for small \mathbf{q} vectors, and therefore we neglect those terms in the following discussion. The divergent term $\Pi^{(23)}(\omega = 0, \mathbf{q})$ of the polarization function then takes the form

$$\Pi^{(23)}(\omega = 0, \mathbf{q}) = 2 \int \frac{d^3k}{(2\pi)^3} \frac{[\mathcal{F}(\mathbf{k}, \mathbf{q})^\dagger]_{23} \mathcal{F}(\mathbf{k}, \mathbf{q})_{32}}{\mathcal{E}_2(\mathbf{k} + \mathbf{q}) - \mathcal{E}_3(\mathbf{k})}, \quad (5.97)$$

where $\mathcal{E}_{2,3}(\mathbf{k}) \equiv \mathcal{E}_{\pm,-}(\mathbf{k})$. This integral can be solved analytically, for example by introducing a suitable coordinate system such that $\mathbf{q} = q\mathbf{e}_z$ and spherical coordinates $\{k, \theta, \phi\}$ for the momentum integral over \mathbf{k} . After substituting $x = \cos \theta$ with corresponding integral measure $dx = -\sin \theta d\theta$ we perform a Taylor expansion of both numerator and denominator for small $q \ll \delta m/t \ll \Delta'/t \ll 1$. As a result, the relevant term under discussion then takes the form

$$\Pi^{(23)}(\omega = 0, \mathbf{q}) = \frac{1}{16\pi^2} (at)^2 |\mathbf{q}|^2 \int_0^\Lambda dk k^2 \int_{-1}^1 dx \frac{x^2 \delta m^2}{[(atk - \Delta')^2 + \delta m^2]^{5/2}}. \quad (5.98)$$

Note that the integration over $x \in [-1, 1]$ simply gives a prefactor of $2/3$, and the remaining integral over $k \in [0, \infty)$ also can be easily performed, resulting in:

$$\Pi^{(23)}(\omega = 0, \mathbf{q}) = \frac{1}{144\pi^2} (at)^2 |\mathbf{q}|^2 \left(\frac{\Delta' + \sqrt{\delta m^2 + \Delta'^2}}{\delta m t^{3/2}} \right)^2. \quad (5.99)$$

Close to the bulk quantum critical point, we may simplify our result for $\delta m/t \ll \Delta'/t \ll 1$, such that we obtain the following result for the full polarization function:

$$\Pi(\omega = 0, \mathbf{q}) \approx \frac{1}{18\pi^2} (at)^2 |\mathbf{q}|^2 \frac{\Delta'^2}{t^3 \delta m^2} = \frac{1}{18\pi^2} \frac{\Delta'^2}{t \delta m^2} (v_F^{\text{bulk}} |\mathbf{q}|)^2, \quad (5.100)$$

where $v_F^{\text{bulk}} = ta$ is the bulk Fermi velocity. We have also taken the factor of 2 into account which arises from the fact that $\Pi(\omega = 0, \mathbf{q})_{23} = \Pi(\omega = 0, \mathbf{q})_{32}$. Note that the prefactor $1/(18\pi^2)$ is in very good agreement with numerical calculations of the full polarization function $\Pi(\omega, \mathbf{q})$ in the low-energy limit. The above result shows that the static polarization function $\Pi(\omega = 0, \mathbf{q} \rightarrow 0)$ diverges quadratically as function of δm . Consequently, this implies that the bulk dielectric constant ϵ diverges as well:

$$\epsilon \propto \text{const.} + \frac{\Delta'^2}{\delta m^2} \rightarrow \infty \quad \text{as} \quad \delta m \rightarrow 0. \quad (5.101)$$

Hence, as anticipated from the beginning, this leads to the avoidance of chiral symmetry breaking due to long-range Coulomb interactions:

$$\alpha = \frac{e^2}{\hbar v_F^{\text{surf}} \epsilon} \propto \frac{\delta m}{\Delta'} \rightarrow 0 \quad \text{as} \quad \delta m \rightarrow 0 \quad (5.102)$$

since $v_F^{\text{surf}} \propto \delta m / \Delta'$ [cf. Eq. (5.83)]. Note that this particular behavior can be traced back to the nearby presence of the bulk states which leads to screening of the long-range Coulomb interaction, similar to our discussion of Coulomb interactions in the Fu-Kane-Mele Hamiltonian (cf. section 5.3).

5.4.3 Discussion

To conclude this section, in the minimal model we find a different route to flat surface bands in 3D topological insulators, where the flatness is induced by “spectral pressure” exerted by the bulk bands. However, upon approaching the bulk critical point $m = m_{c1}$, where the bulk band gap closes, we also find a diverging polarization function, $\Pi(\omega = 0, \mathbf{q}) \propto |\mathbf{q}|^2 / (m - m_{c1})^2$, indicating the appearance of screening due to the intermediate nodal lines^{123,124}. Consequently, the effective interaction strength $\alpha = e^2 / (\hbar \epsilon v_F^{\text{surf}})$ vanishes as we approach the quantum critical point $m = m_{c1}$. Therefore, a spontaneous mass generation by virtue of chiral symmetry breaking due to long-range Coulomb interactions does not occur in our minimal model based on strained 3D HgTe. The same reasoning is expected to apply to a large family of bismuth-based compounds for which similar effective models have been constructed recently. However, since the surface bands become increasingly flat close to the critical point, even local interactions are expected to become relevant which might also create a surface band gap, as we discuss below.

5.5 Spontaneous gap generation due to local interactions

So far we have investigated the mechanism of chiral symmetry breaking on the 2D surfaces of 3D topological insulators due to long-range Coulomb interactions. Specifically, in both the Fu-Kane-Mele model and the minimal model based on strained 3D HgTe we have seen that the surface bands become flat as we approach the bulk quantum critical point, but at the same time the polarization function exhibits a divergence due to either nodal lines in the bulk or an intermediate semimetallic phase, which is a precursor to screening of the long-range interactions. Nevertheless, as the Fermi velocity of the surface Dirac fermions vanishes, the average kinetic energy goes to zero as well:

$$\langle \mathcal{E}_{\text{kin}} \rangle \sim \frac{\hbar v_F}{l} = \hbar v_F n_d^{1/2} \rightarrow 0 \quad \text{as} \quad v_F \rightarrow 0. \quad (5.103)$$

Naively, one would expect that any local perturbation on the surface generates a mass gap for the surface states, because the kinetic energy as the relevant energy scale to compare with vanishes. However, as we are concerned with the interactions of 2D surface states of a 3D topological insulator, the situation is more complex: Close to the critical point the penetration length of the surface states into the bulk of the topological insulator may increase and eventually diverge, so that the surface states are being “absorbed” into the continuum of bulk states, resulting in a vanishing effective interaction strength at the surfaces. To study the effect of local interactions in more detail, we first discuss the mean-field argument for the critical interaction strength. After that, we review the role of local interactions in the presence of flat surface states, as exemplified by the Fu-Kane-Mele model and the minimal HgTe model.

In the first place, to develop a rough estimate for the critical interaction strength U_c of a purely local contact interaction $V(\mathbf{r}) = U\delta(\mathbf{r})$ above which a mass term for the Dirac fermions

is generated, we consider the following Dyson-Schwinger equation (cf. section 5.1):

$$\Delta(\epsilon, \mathbf{p}) = i \int_{-\infty}^{\infty} \frac{d\nu}{2\pi} \int \frac{d^2k}{(2\pi)^2} \frac{V_{2D}(\nu - \epsilon, \mathbf{k} - \mathbf{p}) \Delta(\nu, \mathbf{k})}{\nu^2 - v_F^2 |\mathbf{k}|^2 - \Delta(\nu, \mathbf{k})^2 + i0^+}. \quad (5.104)$$

Here, the Fermi velocity v_F of the surface Dirac fermions is finite as long as we are not at the bulk quantum critical point, where the surface bands become flat. A simple estimate for $\Delta(\epsilon, \mathbf{p})$ can be obtained by the mean-field ansatz $\Delta(\epsilon, \mathbf{p}) = \Delta = \text{const}$ for the purely local interaction $V_{2D}(\nu - \epsilon, \mathbf{k} - \mathbf{p}) = U$. In that case, the above self-consistency equation simplifies to

$$1 = i \int_{-\infty}^{\infty} \frac{d\nu}{2\pi} \int \frac{d^2k}{(2\pi)^2} \frac{U}{\nu^2 - v_F^2 |\mathbf{k}|^2 - \Delta^2 + i0^+} = \int \frac{d^2k}{(2\pi)^2} \frac{U}{2\sqrt{v_F^2 |\mathbf{k}|^2 + \Delta^2}}, \quad (5.105)$$

where we have integrated out the internal frequency ν in the last line. If we further assume that Δ is zero below the critical interaction strength U_c and finite above, we can easily perform the momentum integration. As a result, we obtain

$$1 = \frac{1}{2} \int \frac{d^2k}{(2\pi)^2} \frac{U_c}{v_F |\mathbf{k}|} = \frac{U_c}{4\pi v_F} \Lambda, \quad (5.106)$$

where Λ is the momentum cutoff which specifies the range of validity for the Dirac spectrum. An important consequence of the above mean-field calculation is that the critical interaction strength vanishes linearly as v_F goes to zero:

$$U_c = 4\pi \frac{v_F}{\Lambda} \rightarrow 0 \quad \text{as} \quad v_F \rightarrow 0. \quad (5.107)$$

Therefore, since the critical interaction strength U_c vanishes upon approaching the quantum critical point, we generically expect a transition towards a gapped state, even for arbitrarily small interaction strengths. This argument, however, hinges on the assumption that the 2D surface states are well localized at the individual surfaces and that the matrix elements between the surface states of opposite surfaces vanish. Although that assumption proves valid in the Fu-Kane-Mele model, the quantum phase transition in the minimal model is governed by a diverging penetration length of the surface states, as we discuss below.

To describe the boundary states localized at the surfaces of a 3D topological insulator, we consider the following generic ansatz for the wave function ψ :

$$\psi(\mathbf{k}_{\parallel}, r_{\perp}) \sim \psi_0 l^{-1/2} e^{i\mathbf{k}_{\parallel} \cdot \mathbf{r}_{\parallel}} e^{-r_{\perp}/l}. \quad (5.108)$$

Here, the first part describes a plane wave with the conserved, in-plane surface momentum \mathbf{k}_{\parallel} , while r_{\perp} denotes the position with respect to the surface normal vector, and l is the **penetration length** of the surface state into the bulk. Furthermore, ψ_0 describes the spinor structure of the surface states, and the prefactor $l^{-1/2}$ ensures a proper normalization of the surface state:

$$\int_0^{\infty} dr_{\perp} |\psi(\mathbf{k}_{\parallel}, r_{\perp})|^2 \doteq 1. \quad (5.109)$$

In particular, the above ansatz can be utilized to estimate the effective strength U_{eff} of a local Hubbard-like interaction for the surface states:

$$U_{\text{eff}} = \int_0^{\infty} dr_{\perp} \int_0^{\infty} dr'_{\perp} |\psi(\mathbf{k}_{\parallel}, r_{\perp})|^2 U_0 \delta(r_{\perp} - r'_{\perp}) |\psi(\mathbf{k}_{\parallel}, r'_{\perp})|^2 \quad (5.110a)$$

$$\propto U_0 \int_0^{\infty} dr_{\perp} \frac{|\psi(\mathbf{k}_{\parallel}, r_{\perp})|^4}{l^2} \propto \frac{U_0}{l}. \quad (5.110b)$$

Note that the effective interaction strength is reduced by a factor $1/l$ compared to the bare interaction strength U_0 , which allows to distinguish between two cases depending on the behavior of l as function of the tuning parameter of the quantum phase transition.

In the first case, the penetration length l remains finite upon approaching the bulk quantum critical point and generically we expect a gapped surface state, because the critical interaction strength $U_c \propto v_F^{\text{surf}}$ decreases upon approaching the bulk critical point, while U_{eff} remains finite and eventually becomes larger than U_c . We find an example of that particular scenario in the Fu-Kane-Mele model, where the surface bands become increasingly flat as we decrease the spin-orbit coupling λ_{SO} , *i.e.*, $v_F^{\text{surf}} \propto \lambda_{\text{SO}} \rightarrow 0$, but the penetration length l of the surface states remains small. The reason for the exponential localization of the surface states close to the quantum critical point can be traced back to the topological protection of the surface states by the nodal lines in the bulk of the system. As discussed before, those nodal lines enforce the surface bands to become perfectly flat in a large region of the surface Brillouin zone (cf. Fig. 5.21), and consequently the surface states cannot penetrate into the bulk, because most of their weight is localized at the surfaces. We have also checked numerically for a range of parameters that the penetration length l remains finite, even in the presence of perturbations which break the chiral symmetry associated with the presence of the nodal lines. By virtue of Eq. (5.106) we then find that the critical interaction strength $U_c \propto \lambda_{\text{SO}}$ decreases, so that U_{eff} becomes larger than U_c sufficiently close to the critical point $\lambda_{\text{SO}}/t = 0$, resulting in a surface band gap due to local interactions.

To summarize our results for the Fu-Kane-Mele model, we generically expect a spontaneous gap generation due to *local* interactions as a result of the surface bands becoming flat in a large region of the surface Brillouin zone. On the other hand, long-range Coulomb interactions are screened by the nodal lines in the bulk of the system, leading to a diverging dielectric constant ϵ , which ultimately prevents the effective interaction strength $\alpha \propto 1/(\epsilon v_F^{\text{surf}})$ from becoming large.

The situation is quite different in the second scenario, where the penetration length diverges upon approaching the bulk quantum critical point as $l \sim |m - m_{c1}|^{-1}$, so that the effective interaction $U_{\text{eff}} \propto U_0/l$ goes to zero as well as $v_F \sim |m - m_{c1}| \rightarrow 0$ for $m \rightarrow m_{c1}$. Note that we find a realization of such a diverging penetration length in the minimal HgTe model when we tune the mass parameter m towards the critical value m_{c1} . In this case, it is however not clear if local Hubbard-like interactions can lead to a spontaneous gap generation close to the quantum critical point, because both the kinetic energy $\langle \mathcal{E}_{\text{kin}} \rangle \sim ta |m - m_{c1}|$ and the effective interaction strength $U_{\text{eff}} \sim U_0 |m - m_{c1}|$ decrease simultaneously. To answer that question, we have calculated the layer-resolved spin-spin correlation function $\chi(\mathbf{q})$ for the minimal model of strained 3D HgTe (cf. section 3.6) from which we infer the leading instability of the system close to the critical point:

$$\begin{aligned}
\mathcal{H}(\mathbf{k}) = & \sum_z \Psi_{\mathbf{k}z}^\dagger (m - t \cos(k_x a) - t \cos(k_y a)) \Gamma_0 \Psi_{\mathbf{k}z} \\
& - t \sum_z \Psi_{\mathbf{k}z}^\dagger (\sin(k_x a) \Gamma_1 + \sin(k_y a) \Gamma_2) \Psi_{\mathbf{k}z} \\
& - t \sum_z \left[\Psi_{\mathbf{k}z}^\dagger \left(\frac{\Gamma_0 - i\Gamma_3}{2} \right) \Psi_{\mathbf{k},z-a}^\dagger + \text{H.c.} \right] \\
& + \Delta' \sum_z \Psi_{\mathbf{k}z}^\dagger \Gamma_{04} \Psi_{\mathbf{k}z} + \mathcal{H}_Z.
\end{aligned} \tag{5.111}$$

To be specific, we have numerically computed the spin-spin correlation function for the interactions on the individual orbitals for different values of the in-plane surface momentum \mathbf{q} and

layer indices z, z' :

$$\chi(\mathbf{q}) = \mathcal{N} \int \frac{d^2k}{(2\pi)^2} \sum'_{\alpha, \beta} \frac{\langle \mathbf{k}, \alpha | \mathcal{S}_{\pm}(z) | \mathbf{k} + \mathbf{q}, \beta \rangle \langle \mathbf{k} + \mathbf{q}, \beta | \mathcal{S}_{\pm}(z') | \mathbf{k}, \alpha \rangle + \text{c.c.}}{\mathcal{E}_{\mathbf{k}, \alpha} - \mathcal{E}_{\mathbf{k} + \mathbf{q}, \beta}}, \quad (5.112)$$

Here, Σ' indicates that the summation is restricted to states which fulfill the energy condition $\mathcal{E}_{\mathbf{k}, \alpha} < \mathcal{E}_F < \mathcal{E}_{\mathbf{k} + \mathbf{q}, \beta}$, and \mathcal{N} is a normalization constant. Furthermore, the spin-operator $\mathcal{S}_{\pm}(z)$ acting in layer z is defined by:

$$\mathcal{S}_{\pm}(z) \equiv \left(\frac{\tau_0 \pm \tau_z}{2} \right) \otimes \sigma_z, \quad (5.113)$$

where the “ \pm ” determines whether we compute the spin-spin correlation function and thus the leading instability of the electron-like (E) or light-hole (LH) states (for a definition of the states involved see section 3.6).

Fig. 5.27 shows the eigenvalues of the susceptibility matrix $\chi(\mathbf{q})$ for a slab with a thickness of $N_z = 20$ layers and for different high-symmetry points in the surface Brillouin zone. Upon approaching the bulk quantum critical point $m \rightarrow m_{c1}$, where roughly $m_{c1} \approx 2.6 - 2.7$, the leading surface instability close to the critical point is found to be of ferromagnetic type, as indicated by the open red squares in Fig. 5.27. To be precise, we find the following order of eigenvalues of $\chi(\mathbf{q})$:

$$\mathcal{E}_{(0,0,0)}^{\text{bulk}} < \mathcal{E}_{(0,0,\pi)}^{\text{bulk}} < \mathcal{E}_{(0,0)}^{\text{surf}}, \quad (5.114a)$$

$$\mathcal{E}_{(\pi,0)}^{\text{surf}} < \mathcal{E}_{(\pi,0,\pi)}^{\text{bulk}} < \mathcal{E}_{(\pi,0,0)}^{\text{bulk}}, \quad (5.114b)$$

$$\mathcal{E}_{(\pi,\pi)}^{\text{surf}} < \mathcal{E}_{(\pi,\pi,\pi)}^{\text{bulk}} < \mathcal{E}_{(\pi,\pi,0)}^{\text{bulk}}. \quad (5.114c)$$

Note that none of the eigenvalues of the susceptibility matrix $\chi(\mathbf{q})$ diverges upon approaching the critical point m_{c1} , so that the usual mean-field condition $\mathbb{1} \doteq U\chi(\mathbf{q})$ implies a finite interaction strength U^* for a gapped surface state. However, since the eigenvalues become smaller close to m_{c1} , the critical interaction strength U^* also becomes larger, but does not diverge. Furthermore, note that finite-size effects are not small for this system size and lead to corrections of the bulk eigenvalues. For example, although the two momenta $(\pi, 0, \pi)^T$ and $(\pi, \pi, 0)^T$ are equivalent by virtue of lattice symmetries, their eigenvalues differ by up to 10%. We have checked numerically that those deviations become smaller when increasing the slab width to larger values, and we have also analyzed the structure of the eigenvectors as function of the layer index z for the different surface momenta (not shown here). The above results are summarized in the schematic plot of the eigenvalues shown in Fig. 5.28, where bulk eigenvalues corresponding to symmetry-equivalent points in the Brillouin zone have the same value, and the surface eigenvalues are indicated by open symbols, showing that the leading instability is surface ferromagnetism.

Fig. 5.29 shows the surface and bulk eigenvectors of the susceptibility matrix $\chi(\mathbf{q} \rightarrow 0)$ for a slab with $N_z = 20$ layers and tuning parameter $m/t = 2$ [cf. Fig. 5.27 (a)]. The two eigenvectors of the doubly degenerate surface eigenvalue are well localized at the surfaces of the slab (open symbols), as shown in Fig. 5.29 (a), where most of their weight is localized in the first three layers of the slab. On the other hand, the eigenvector corresponding to the largest bulk eigenvalue (filled green symbols) shows a rapidly oscillating behavior as a function of the layer index which can be reconciled with a momentum $q_z = \pi/a$ ($a = 1$) perpendicular to the surface [cf. Fig. 5.29 (b)]. In that sense, we may identify this bulk eigenvalue with the bulk momentum $(0, 0, \pi)^T$. Similarly, the eigenvectors of the two smallest bulk eigenvalues shown in Fig. 5.29 (b)

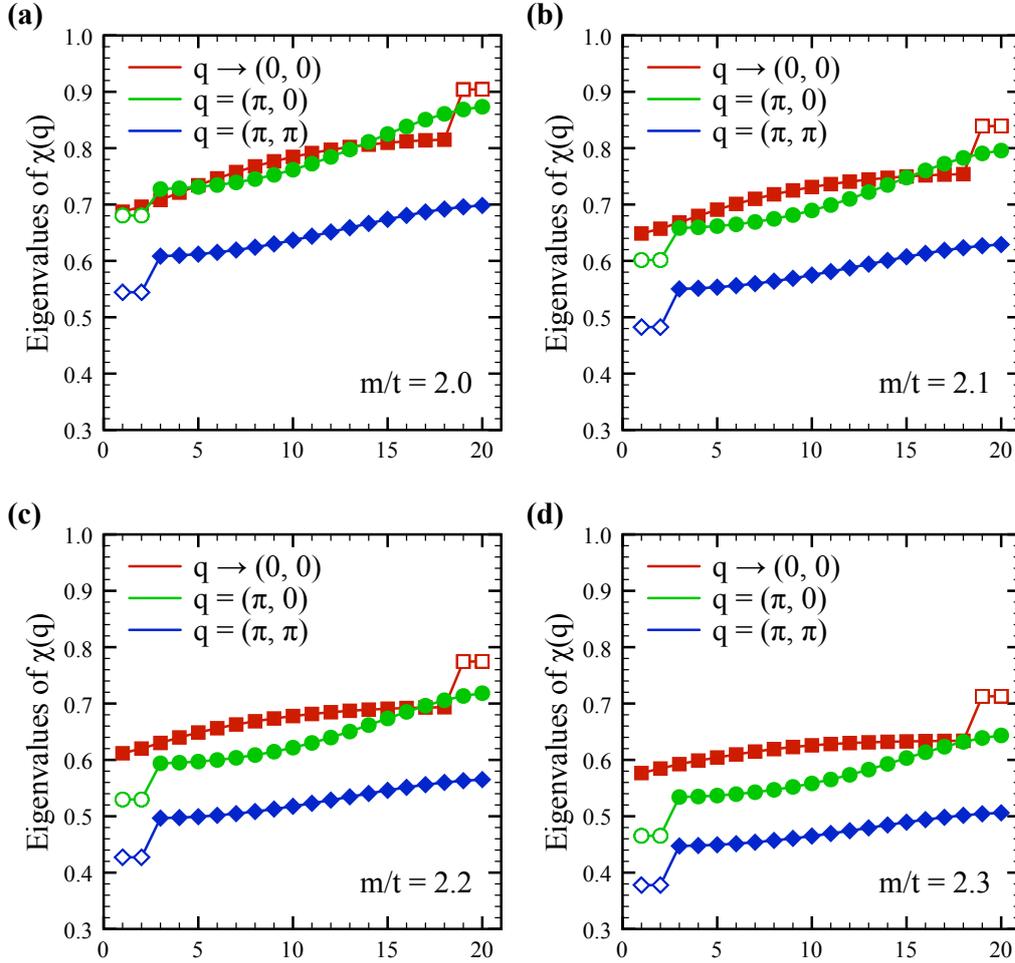


Figure 5.27. Eigenvalues of the layer-resolved susceptibility matrix $\chi(\mathbf{q})$ for the light-hole orbitals [cf. Eq. (5.112) and discussion in main text] for a slab with a thickness of $N_z = 20$ layers and for three high-symmetry surface momenta $\mathbf{q} = (0.01, 0.01)^T$, $\mathbf{q} = (\pi, 0)^T$, and $\mathbf{q} = (\pi, \pi)^T$ indicated in red, green, and blue. The open and filled symbols denote the eigenvalues, whose eigenvectors have the most weight at one of the surfaces or in the bulk, respectively. The parameters of the minimal HgTe model introduced in chapter 3 have been chosen as $t = 1$, $\Delta'/t = 0.7$, realizing an inversion-asymmetric strong topological insulator for $m < m_{c1}$, where roughly $m_{c1} \approx 2.6 - 2.7$. For small slabs, finite-size effects lead to corrections of the eigenvalues, so that, for example, the bulk eigenvalues corresponding to the momenta $(\pi, 0, \pi)^T$ (smallest bulk eigenvalue for $\mathbf{q} = (\pi, 0)^T$) and $(\pi, \pi, 0)^T$ (largest eigenvalue for $\mathbf{q} = (\pi, \pi)^T$) do not have the same value, but differ by up to 10%. Those deviations, however, become smaller when increasing the slab thickness N_z .

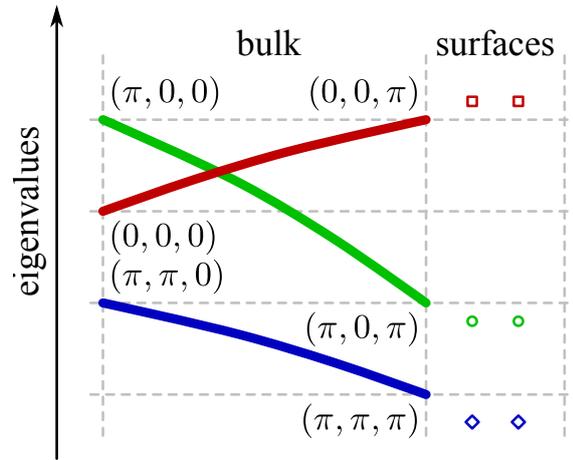


Figure 5.28. Schematic plot of the eigenvalues of the susceptibility matrix $\chi(\mathbf{q})$ [cf. Eq. (5.112)] for three high-symmetry surface momenta $\mathbf{q} = (0, 0)^T$, $\mathbf{q} = (\pi, 0)^T$, and $\mathbf{q} = (\pi, \pi)^T$ indicated in red, green, and blue. The thick lines represent those eigenvalues, whose eigenvectors are localized in the bulk of the system, while the open symbols represent the eigenvalues of the eigenvectors located at the surfaces of the slab (see Fig. 5.27). Importantly, the largest eigenvalue is found for $\mathbf{q} = (0, 0)^T$ at the surfaces, implying that the leading instability of the system is a ferromagnetic one, while the largest bulk eigenvalue with momentum $(0, 0, \pi)^T$ shows an anti-ferromagnetic order in the direction normal to the surfaces. Note that in large systems the bulk eigenvalues corresponding to equivalent momenta, e.g., $(\pi, 0, 0)^T$ and $(0, 0, \pi)^T$ have the same value.

vary smoothly across the slab, corresponding to small bulk momenta $\mathbf{q} \rightarrow (0, 0, 0)^T$. We have also checked numerically that the eigenvectors of bulk eigenvalues for different surface momenta are equivalent up to symmetry transformations of the underlying simple cubic lattice. Combining those results, we obtain the first set of inequalities in Eq. (5.114), and a similar analysis of the eigenmodes of $\chi(\mathbf{q})$ for varying momenta then gives rise to the schematic plot shown in Fig. 5.28.

To summarize our results regarding local interactions in the minimal model based on HgTe, the leading instability of the 2D surfaces is a ferromagnetic one, with an anti-ferromagnetic coupling of the two surfaces, as indicated by the largest bulk eigenvalue. However, since all eigenvalues of the susceptibility matrix $\chi(\mathbf{q})$ are of the same order, this instability only occurs in a very small region of the m -vs.- U phase diagram, so that fine-tuning of the interaction strength U and the mass parameter m of the model is required. In contrast to the Fu-Kane-Mele model, the spontaneous gap generation of the surface Dirac fermions in the minimal model is not expected to be a generic feature, but it rather depends on microscopic details of the model.

5.6 Discussion and outlook

Finally, let us put our results for the mechanism of a spontaneous mass generation on the surfaces of 3D topological insulators in a broader context following the theoretical classification of quantum phase transitions of three-dimensional, time-reversal invariant topological insulators due to Murakami^{123,124}. In that classification scheme, the quantum phase transition from a topologically non-trivial state to a band insulator state as function of a single tuning parameter m is shown to depend on the presence or absence of inversion symmetry, giving rise to two classes \mathcal{A} and \mathcal{B} of Hamiltonians, respectively. The questions we would like to address in the following are how the surface states and correspondingly the topological \mathbb{Z}_2 invariant ν_0 of

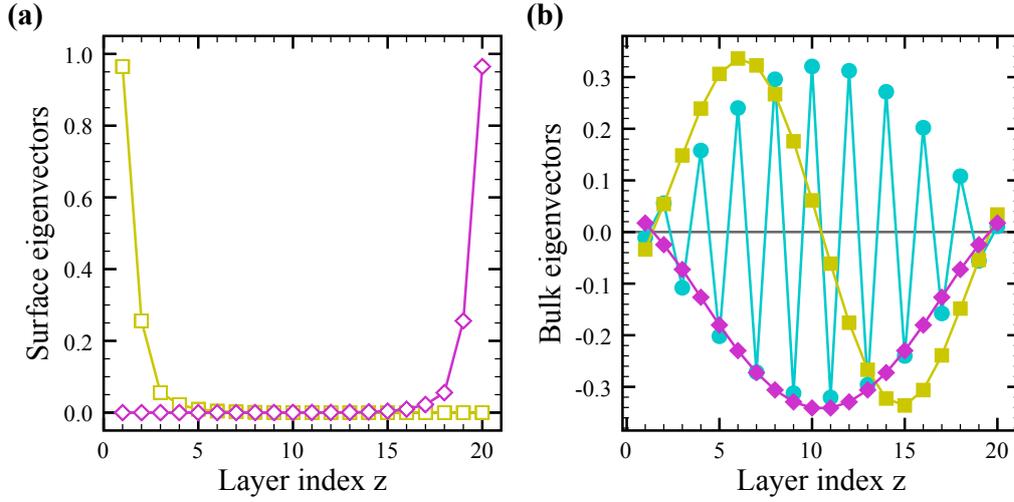


Figure 5.29. Eigenvectors of the layer-resolved susceptibility matrix $\chi(\mathbf{q} \approx 0)$ for a slab thickness of $N_z = 20$ layers, and the parameters of the minimal HgTe model are the same as in Fig. 5.29. **(a)** Localization of the two surface eigenvectors indicated by open symbols, where most of the weight is located within the first three layers of the slab. **(b)** The eigenvector of the largest bulk eigenvalue indicated by filled circles shows a rapidly oscillating behavior as function of the layer index z which can be associated with a momentum of $q_z = \pi$ perpendicular to the slab surfaces. On the other hand, the eigenvectors of the smallest bulk eigenvalues indicated by filled squares and filled diamonds show a smooth variation across the slab which is indicative of small momenta.

a 3D topological insulators changes when we tune an external parameter, and whether or not we can expect a spontaneous mass generation for the surface Dirac fermions due to long-range Coulomb interactions or short-ranged, Hubbard-like interactions. Such a system with a finite surface band gap is a prerequisite for more exotic states of topological insulators showing the anomalous quantum Hall effect or the magneto-electric effect, as mentioned before.

In the following discussion, we consider “generic” quantum phase transitions between topologically trivial and non-trivial phases as a function of a *single* tuning parameter m , and we exclude phase transitions where a crossing of bulk bands is achieved by tuning two or more parameters. Similar to our approach of tuning the spin-orbit coupling λ_{SO} discussed above, this generic parameter m can be thought of as being externally controllable, for example by exerting pressure or by chemical doping. Note that at the quantum phase transition, the topological \mathbb{Z}_2 invariant ν_0 must change its character from being odd to being even, or vice versa. According to the bulk-boundary correspondence, this implies that the bulk band gap must close at some point \mathbf{k}^* in the Brillouin zone, because only then can the bulk bands “unwind.” To study that quantum phase transition, let us consider the following generic Bloch Hamiltonian:

$$\mathcal{H}(\mathbf{k}) = \begin{bmatrix} h_{\uparrow\uparrow}(\mathbf{k}) & h_{\uparrow\downarrow}(\mathbf{k}) \\ h_{\downarrow\uparrow}(\mathbf{k}) & h_{\downarrow\downarrow}(\mathbf{k}) \end{bmatrix}. \quad (5.115)$$

If we assume that the Hamiltonian has no additional degeneracies other than those imposed by time-reversal symmetry, it suffices to discuss the phase transition of 3D topological insulators in terms of 4×4 Hamiltonian matrices, because we need two spin states and two bands with opposite parity which are exchanged at the phase transition. **Time-reversal symmetry** relates the spin-up and spin-down parts of the Bloch Hamiltonian at momenta \mathbf{k} and $-\mathbf{k}$ to each other

as follows:

$$\mathcal{H}(\mathbf{k}) = \hat{\Theta}\mathcal{H}(-\mathbf{k})^T\hat{\Theta}^{-1}, \quad (5.116)$$

where $\hat{\Theta} = i\sigma_y \otimes \tau_0 \mathcal{K}$ implements the time-reversal operator. As introduced before, $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)^T$ acts on the spin degrees of freedom, whereas $\boldsymbol{\tau} = (\tau_x, \tau_y, \tau_z)^T$ acts on the orbital degrees of freedom, and $\sigma_0 = \tau_0 = \mathbf{1}$ is the 2×2 identity matrix. In particular, Eq. (5.116) implies that

$$\mathfrak{h}_{\sigma\sigma}(\mathbf{k}) = +\mathfrak{h}_{\bar{\sigma}\bar{\sigma}}(-\mathbf{k})^T, \quad (5.117a)$$

$$\mathfrak{h}_{\sigma\bar{\sigma}}(\mathbf{k}) = -\mathfrak{h}_{\sigma\bar{\sigma}}(-\mathbf{k})^T, \quad (5.117b)$$

where $\sigma \in \{\uparrow, \downarrow\}$ and $\bar{\sigma} \in \{\downarrow, \uparrow\}$ denote opposite spin orientations. From the Kramers' theorem for time-reversal invariant systems it follows that the spectrum of the Bloch Hamiltonian $\mathcal{H}(\mathbf{k})$ is doubly degenerate at the eight time-reversal invariant momenta Γ_i ($i = 1, \dots, 8$) which can be parametrized as

$$\Gamma_i = \Gamma_{(n_1 n_2 n_3)} = \frac{1}{2} \sum_{l=1}^3 n_l \mathbf{b}_l, \quad (5.118)$$

where $\mathbf{b}_1, \mathbf{b}_2,$ and \mathbf{b}_3 denote the primitive reciprocal lattice vectors of the system. There is, however, another important symmetry which relates the spectrum at opposite momenta \mathbf{k} and $-\mathbf{k}$, namely **inversion symmetry** $\hat{\mathcal{P}}$:

$$\mathcal{H}(\mathbf{k}) = \hat{\mathcal{P}}\mathcal{H}(-\mathbf{k})\hat{\mathcal{P}}^{-1}. \quad (5.119)$$

The presence or absence of inversion symmetry has important consequences on the existence of a non-trivial, semimetallic phase between the quantum spin Hall insulator and the trivial band insulator, as we discuss below. Since other kinds of band crossings due to higher point-group symmetries can be considered as degenerate cases of the generic system discussed in this section, we only consider the effects of time-reversal and inversion symmetry in the following.

5.6.1 Inversion-symmetric systems

In the first place, let us discuss systems *with* inversion symmetry. As a consequence of both time-reversal invariance and inversion symmetry, the bulk bands are doubly degenerate for all momenta \mathbf{k} due to Kramers' theorem, and the quantum phase transition occurs when the band gap between the two doubly-degenerate conduction and valence bands closes at some point \mathbf{k}^* in the Brillouin zone. Furthermore, in the presence of inversion symmetry, the two bands involved can be characterized by their parity eigenvalues $\eta_1, \eta_2 = \pm 1$. If both bands have the same parity, $\eta_1 = \eta_2$, then the co-dimension, *i.e.*, the number of conditions which have to be fulfilled simultaneously for the band gap to close, is five^{123,124}. In particular, this exceeds the number of tunable parameters $m, k_x, k_y,$ and k_z . Hence, in that case there is no generic band crossing at any point \mathbf{k}^* in the Brillouin zone.

The situation is, however, different if the two bands have opposite parity, *i.e.*, we consider $\eta_1 = -\eta_2$. In that case, the Bloch Hamiltonian can be parametrized as

$$\mathcal{H}(\mathbf{k}) = \epsilon(\mathbf{k}) + d_0(\mathbf{k})\Gamma_0 + \sum_{j=1}^4 d_j(\mathbf{k})\Gamma_j, \quad (5.120)$$

where $\epsilon(\mathbf{k})$ and $d_0(\mathbf{k})$ are even functions of \mathbf{k} , while $d_j(\mathbf{k})$ ($j = 1, \dots, 4$) are odd functions of \mathbf{k} . A suitable set of matrices Γ_i ($i = 0, \dots, 4$) has been introduced in chapter 3 for the minimal

model of a topological insulator based on strained 3D HgTe, where the two bands involved correspond to the s -like Γ_6 band with even parity and the p -like Γ_8 band with odd parity, but this representation of $\mathcal{H}(\mathbf{k})$ is, of course, not limited to a specific system. Since the above Γ matrices anti-commute as discussed in section 3.6.6, the electronic band structure of the model Hamiltonian is simply given by:

$$\mathcal{E}(\mathbf{k}) = \pm |\mathbf{d}(\mathbf{k})| = \pm \sqrt{\sum_{j=0}^4 d_j(\mathbf{k})^2}, \quad (5.121)$$

where each band is doubly degenerate, as mentioned before. Although we cannot expect a band crossing for a generic point \mathbf{k}^* in the Brillouin zone as the co-dimension is also five, the number of tunable parameters reduces to just m at the time-reversal invariant momenta Γ_i , because $d_j(\mathbf{k} = \Gamma_i) = 0$ (for all $j = 1, \dots, 4$) by virtue of being odd functions of \mathbf{k} with respect to Γ_i . Hence, only for bands with opposite parities we expect a crossing of the doubly-degenerate bands at some $\mathbf{k}^* = \Gamma_i$. On one side of the quantum phase transition, *e.g.*, for $m < m_0$, the system describes a topologically non-trivial phase like the quantum spin Hall insulator, while the other side $m > m_0$ describes the topologically trivial band insulator.

Thus, in time-reversal invariant systems with two bands of opposite parity and an additional inversion symmetry, it is possible to tune a quantum phase transition between different topological sectors by changing a *single* parameter m . In particular, we have discussed the Fu-Kane-Mele model on the diamond lattice, where the phase transition of the corresponding strong topological insulator phase to the trivial phase can be controlled by changing the spin-orbit coupling parameter λ_{SO}/t . However, note that to realize a strong topological insulator phase in the Fu-Kane-Mele model it is necessary to include anisotropic modulations of the overlap parameter which effectively shift the crossing points away from the Γ_i into the bulk Brillouin zone (cf. section 5.3 for details).

5.6.2 Inversion-asymmetric systems

In systems *without* inversion symmetry, it is generically impossible to observe the quantum phase transition between topologically distinct systems for a single value of the tuning parameter m , as we discuss now. Let us first consider the time-reversal invariant momenta Γ_i , where the spectrum is doubly degenerate due to Kramers' theorem. Since the 4×4 Bloch Hamiltonian can be written in a form similar to Eq. (5.120), the doubly degenerate conduction and valence bands touch only when all $d_j(\mathbf{k})$ vanish, implying that the co-dimension is equal to five^{123,124}. However, the momentum $\mathbf{k} = \Gamma_i$ is fixed, so that one cannot control the band touching at the time-reversal invariant momenta Γ_i by tuning the single remaining parameter m .

Away from the time-reversal invariant momenta, *i.e.*, for $\mathbf{k} \neq \Gamma_i$, the spectrum is non-degenerate, and the gap closes due to a touching of only two bands. Such a band crossing can be described in terms of an effective 2×2 Bloch Hamiltonian:

$$\mathcal{H}_{\text{eff}}(\mathbf{k}) = \sum_{j=0}^3 \tilde{d}_j(\mathbf{k}) \sigma_j = \begin{bmatrix} \tilde{d}_0(\mathbf{k}) + \tilde{d}_3(\mathbf{k}) & \tilde{d}_1(\mathbf{k}) - i\tilde{d}_2(\mathbf{k}) \\ \tilde{d}_1(\mathbf{k}) + i\tilde{d}_2(\mathbf{k}) & \tilde{d}_0(\mathbf{k}) - \tilde{d}_3(\mathbf{k}) \end{bmatrix}, \quad (5.122)$$

where $\tilde{d}_j(\mathbf{k})$ ($j = 0, \dots, 3$) are real functions of m and \mathbf{k} . The two bands touch if and only if the following three conditions are met:

$$\tilde{d}_j(\mathbf{k}) \doteq 0 \quad \forall j \in \{1, 2, 3\}, \quad (5.123)$$

because $\tilde{d}_0(\mathbf{k})$ only leads to a shift of the energy levels, but does not affect the level crossing otherwise. Notably, these three conditions define a one-dimensional curve in the four-dimensional (m, k_x, k_y, k_z) space. Hence, for a generic value of the tuning parameter m there will be points \mathbf{k}^* , where the bands cross, and when m is changed adiabatically, those points move in momentum space, so that the system remains gapless. Because the two sides of the quantum phase transition ($m \ll m_0$ and $m \gg m_0$) belong to topologically distinct phases, there should be some phase transition in between, and one can thus ask how this phase transition looks like in the inversion-asymmetric systems upon changing the tuning parameter m .

The answer to that question is that instead of a phase transition at a *single* value m_0 of the tuning parameter there appears a finite *region* $m_{c1} < m < m_{c2}$, where the system remains gapless. Furthermore, the only way to open or close the bulk band gap is to create or destroy pairs of gapless points or, more precisely, pairs of so-called **Weyl fermions**. Such gapless points with linear spectrum represent monopoles in momentum space and carry a topological quantum number or “charge”^{58,180–182}. For example, a pair of a monopole with charge $q = 1$ and an anti-monopole with charge $q = -1$ is created at $m = m_{c1}$, so that the system becomes gapless. By further increasing the tuning parameter m , the monopole and anti-monopole at $\mathbf{k}_\pm(m)$ start to shift around in momentum space as function of m , but time-reversal symmetry always enforces that $\mathbf{k}_+ = \mathbf{k}_-$. Importantly, a bulk band gap can only open again when the monopole and the anti-monopole meet and annihilate which happens at $m = m_{c2}$, and the system enters the trivial band insulating phase.

5.6.3 Overview of the quantum phase transitions of 3D topological insulators

To summarize, the different quantum phase transitions of time-reversal invariant 3D topological insulators with and without inversion symmetry are schematically depicted in Fig. 5.30, while Fig. 5.31 shows the corresponding phase diagram in the plane of a single tuning parameter m and a second parameter δ describing an inversion-symmetry breaking perturbation of the system. Fig. 5.30 (a) shows the phase transition of an inversion-symmetric topological insulator occurring at a single point m_0 in the parameter space. On the other hand, a topologically non-trivial gapless phase $m_{c1} < m < m_{c2}$ with Weyl fermions appears between the quantum spin Hall insulator and the trivial band insulator when inversion symmetry is broken [see Fig. 5.30 (b)], giving rise to the intermediate gapless region in the phase diagram in Fig. 5.31.

At this point, one might ask the question how inversion symmetry modifies the structure of the surface bands and whether this symmetry is important in the context of a spontaneous gap generation due to long-range Coulomb interactions. Furthermore, the intermediate semimetallic Weyl phase could also affect the screening of Coulomb interactions, so that a spontaneous symmetry breaking due to long-range interaction might be prevented. To answer that question one may discuss different scenarios for gapped surface states which may arise in inversion-symmetric and inversion-asymmetric systems.

5.6.4 Flat surface bands and spontaneous mass generation

Within the above classification of quantum phase transitions of time-reversal invariant 3D topological insulators we may discuss possible scenarios for the spontaneous gap generation of the surface states. As discussed before, one would in general expect that a flat surface band leads to a diverging effective interaction strength as $\alpha \propto 1/(\epsilon v_F^{\text{surf}})$, and thereby induces a spontaneous surface band gap by means of chiral symmetry breaking due to long-range Coulomb interactions if we are able to tune the surface Fermi velocity v_F^{surf} appropriately (cf. section 5.1). However, close to the critical point screening of the long-range interactions becomes more and

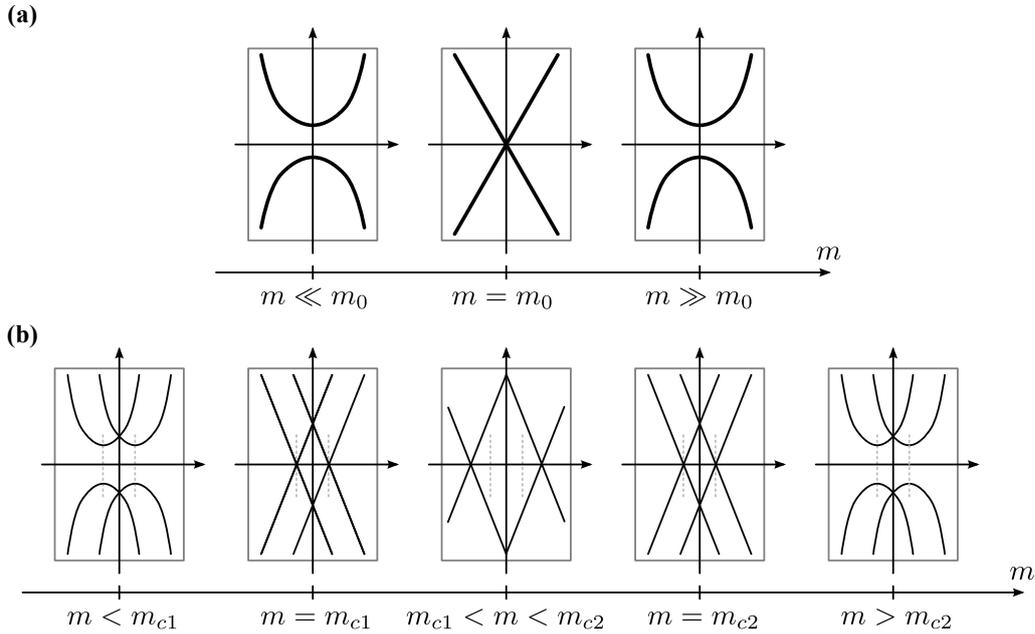


Figure 5.30. Schematic illustration of the quantum phase transition of a time-reversal invariant 3D topological insulator in the (a) inversion-symmetric and (b) inversion-asymmetric case as a function of a single tuning parameter m . The left- and right-most panels indicate the quantum spin Hall insulator for $m \ll m_0$ and the trivial band insulator for $m \gg m_0$, respectively. In inversion-symmetric systems, the quantum phase transition between the two distinct phases occurs at a single value m_0 of the tuning parameter, while in systems without inversion symmetry there appears a gapless phase for $m_{c1} < m < m_{c2}$. In this intermediate phase, Weyl fermions with a linear spectrum shift around in the Brillouin zone as function of the tuning parameter m until they meet again at $m = m_{c2}$, where they annihilate each other and create a bulk band gap again. Figure after Ref. 123.

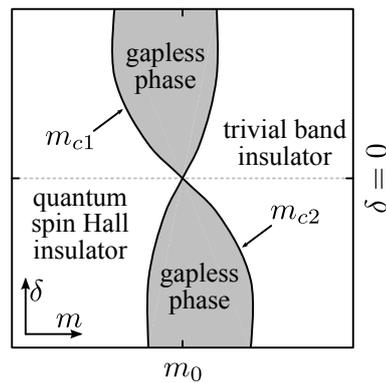


Figure 5.31. Schematic illustration of the generic phase diagram of time-reversal invariant 3D topological insulators. m denotes the tuning parameter, while δ describes an inversion-symmetry breaking perturbation of the system, and $\delta = 0$ describes inversion-symmetric systems. The boundaries of the intermediate, gapless phase of Weyl fermions determine the critical values m_{c1} and m_{c2} of the tuning parameter. Figure after Ref. 123.

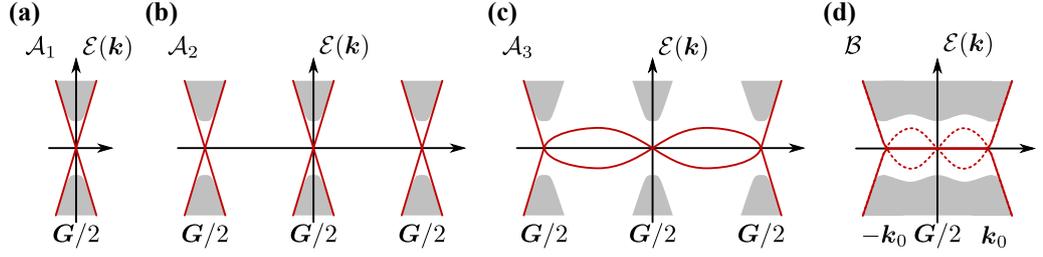


Figure 5.32. Schematic illustration of the different classes \mathcal{A}_1 , \mathcal{A}_2 , \mathcal{A}_3 , and \mathcal{B} within Murakami's classification of phase transitions of 3D topological insulators^{123,124}. (a) In class \mathcal{A}_1 , there exists only a single bulk Dirac point (indicated by gray shaded area) at a time-reversal invariant momentum $G/2$, and the corresponding surface bands (indicated by red lines) inherit its Fermi velocity, *i.e.*, we get $v_F^{\text{surf}} \propto v_F^{\text{bulk}}$. (b) In class \mathcal{A}_2 , there exists more than one bulk Dirac point, but they remain disconnected from each other. Hence, the corresponding surface bands is determined by the bulk Fermi velocity, *i.e.*, we again have $v_F^{\text{surf}} \propto v_F^{\text{bulk}}$. (c) If, however, the bulk Dirac points become connected by surface bands, this opens up the possibility to control the Fermi velocity of the surface Dirac fermions independently of the bulk Fermi velocity. (d) In inversion-asymmetric systems of class \mathcal{B} , the Dirac points are located symmetrically around the time-reversal invariant momenta $k = G/2 \pm k_0$. Similar to class \mathcal{A}_3 the Fermi velocity of the surface bands connecting pairs of Dirac points can be reduced upon approaching the bulk quantum critical point, because the bulk conduction and valence bands exert “spectral pressure” onto the surface bands.

more relevant, and in both the Fu-Kane-Mele model and the minimal HgTe model we find that the dielectric constant ϵ diverges upon approaching the bulk critical point. Despite the fact that the surface bands become increasingly flat, long-range Coulomb interactions are in general not expected to induce a surface band gap. On the other hand, local interactions can also lead to a band gap if the average kinetic energy as the relevant energy scale to compare with becomes smaller and smaller close to the critical point. In that setup, the Fu-Kane-Mele model shows a generic instability towards a ferromagnetic surface as the surface states become flat in a large region of the corresponding surface Brillouin zone. The minimal HgTe model on the other hand only shows a ferromagnetic surface instability in a small region of the phase diagram which requires fine-tuning of the model parameters, and in that model it comes down to numbers.

To develop a better understanding of those results in a broader context, let us first consider strong topological insulators with \mathbb{Z}_2 invariant $\nu_0 = 1$ in the presence of inversion symmetry (class \mathcal{A}), where the bulk band gap closes at one of time-reversal invariant momenta in the Brillouin zone. Since strong topological insulators are characterized by an odd number of Dirac cones in the corresponding surface Brillouin zone, we may distinguish between three subclasses [cf. Fig. 5.32 (a-c)].

In class \mathcal{A}_1 , there exists only a single Dirac point located at one of the time-reversal invariant momenta in the Brillouin zone, $k_0 = \Gamma_i$. Since the surface bands connect the bulk conduction and valence bands across k_0 , their Fermi velocity is mainly determined by the bulk Fermi velocity, *i.e.*, we find $v_F^{\text{surf}} \propto v_F^{\text{bulk}}$ [cf. Fig. 5.32 (a)], and we cannot tune the surface Fermi velocity v_F^{surf} by varying the tuning parameter m of the model. As a consequence, we have no control over the effective interaction strength $\alpha = e^2 / (\hbar \epsilon v_F^{\text{surf}}) \propto 1 / (\epsilon v_F^{\text{surf}})$, so that the spontaneous generation of a mass gap due to long-range Coulomb interactions upon approaching the bulk quantum critical point is not expected to be a generic feature in class \mathcal{A}_1 . Furthermore, as the average kinetic energy of the surface Dirac fermions remains finite, we cannot hope to find a generic mass gap due to local interactions either.

An interesting situation, however, might be realized in a certain class of 3D “frustrated hop-

ping” models such as the non-magnetic pyrochlore systems with a single Dirac fermion on the surfaces. In those systems, band structure calculations have shown that one can find degenerate, flat bulk bands in the absence of spin-orbit interactions⁵⁶. This, in principle, would allow to investigate the mechanism of spontaneous gap generation also in such a system by varying the spin-orbit coupling strength. Whether a particular pyrochlore system is a strong topological insulator can, however, only be established by means of detailed band structure calculations or experiments. Although long-range Coulomb interactions are expected to be screened due to the large density of states induced by the flat bands, local interactions could in general induce a surface band gap, because the average kinetic energy of the Dirac fermions becomes small close to the quantum critical point. Consequently, those systems should present promising models for spontaneously gapped surface states which would be worth studying in the future.

Two further subclasses \mathcal{A}_2 and \mathcal{A}_3 of inversion-symmetric topological insulators exist, in which an odd number $n > 1$ of Dirac points located at the time-reversal invariant momenta occur. To start with, let us consider class \mathcal{A}_2 , where the Dirac points are independent of each other in the sense that they are not connected by surface bands [see Fig. 5.32 (b)]. Here, the Fermi velocity of the surface Dirac fermions is again determined by the bulk Fermi velocity, $v_F^{\text{surf}} \propto v_F^{\text{bulk}}$, and similar to class \mathcal{A}_1 we do not generically expect a spontaneously generated mass on those surfaces either.

If, however, the Dirac points are indeed connected by the surface bands as in class \mathcal{A}_3 [see Fig. 5.32 (c)], we can, in principle, hope to drive the system towards a critical point, where the surface bands become increasingly flat, so that a band gap for the surface Dirac fermions could be spontaneously generated by the mechanism of chiral symmetry breaking due to long-range Coulomb interactions. One particular example of such a system in class \mathcal{A}_3 is the Fu-Kane-Mele model on the 3D diamond lattice, where we have investigated the flatness of the surface states and where we have shown that we get perfectly flat surface bands with $v_F^{\text{surf}}/t \rightarrow 0$ upon approaching the bulk quantum critical point. Here, the existence and topological stability of those zero-energy surface bands is guaranteed by an additional symmetry of the model, namely chiral symmetry, which leads to nodal lines in the bulk of the system, *i.e.*, one-dimensional curves of Dirac points with zero energy (cf. section 5.3.4). In particular, since those nodal lines have finite components in all spatial directions, they are also responsible for the screening of long-range Coulomb interactions which manifests itself in a diverging dielectric constant (cf. section 5.3.5). Hence, in a generic system, we do not expect that one can tune the effective interaction strength $\alpha \propto 1/(\epsilon v_F^{\text{surf}})$ to values larger than the critical interaction α_c , where a spontaneous generation of a surface band gap occurs.

In the presence of spin-orbit interactions λ_{SO} or hopping between second-nearest neighbors t' we also do not expect that Coulomb interactions can induce a spontaneous mass in the surface Dirac fermions, since both interactions break the symmetry of the Fu-Kane-Mele model. To be precise, spin-orbit interactions lift the degeneracy of the nodal bulk lines, so that only a finite number of bulk Dirac points remains which are subsequently gapped if we also take modulated hopping parameters δt_j into account. Hopping between second-nearest neighbors, however, only leads to an energy shift of the nodal lines which can be seen from the corresponding density of states (cf. section 5.3.2). In particular, the surface bands no longer remain perfectly flat in the presence of those interactions, because longer-ranged hopping processes lead to a finite kinetic energy of the surface states. Hence, we expect to find surface states with a small Fermi velocity, but being close to the bulk quantum critical point the dielectric constant ϵ is expected to become large as well as a precursor to screening. As a consequence, chiral symmetry breaking due to long-range Coulomb interactions does not play a role in the Fu-Kane-Mele model. However, as the average kinetic energy becomes smaller and smaller upon approach-

ing the bulk quantum critical point, local interactions are expected to create a gap, discussed in section 5.5. In particular, the localization of the surface states within the penetration length l leads to a finite effective interaction strength $U_{\text{eff}} \sim U_0/l$, where U_0 is the bare, local Hubbard interaction. As the critical interaction strength $U_c \sim v_F$ decreases, it will eventually become smaller than U_{eff} , and quite generically a surface band gap develops.

However, it is not clear whether the presence of a chiral symmetry as in the Fu-Kane-Mele model is the generic situation in class \mathcal{A}_3 to achieve flat surface state bands. One could also try to construct certain toy models, where hopping in one direction is different than in the two orthogonal directions such that bulk nodal lines (provided they exist) will be confined to two-dimensional planes in momentum space. Since those nodal lines will not extend in all momentum directions one might expect anisotropic screening of long-ranged Coulomb interactions perpendicular and parallel to the nodal lines. Consequently, the polarization function will diverge only in, e.g., the q_x and q_y direction, but show a smooth behavior as function of q_z . Such an anisotropic screening might open up the possibility for a “partial” chiral symmetry breaking due to Coulomb interactions on certain surfaces.

In the second class of quantum phase transitions between the topologically distinct phases, which occurs in systems with broken inversion symmetry (class \mathcal{B}), one would expect a better situation for the spontaneous gap generation as far as the generic appearance of flat surface bands is concerned: First, the surface bands connect pairs of bulk Dirac points which are located symmetrically at $\mathbf{k} = \mathbf{G}/2 \pm \mathbf{k}_0$ around the time-reversal invariant momenta of the Brillouin zone [see Fig. 5.32 (d)]. Due to the finite distance of those Dirac points set by $|\mathbf{k}_0|$, we can give a rough estimate for the Fermi velocity of the surface Dirac fermions as $v_F^{\text{surf}} \propto \mathcal{E}_{\text{gap}}/|\mathbf{k}_0|$. As we have shown in the minimal model derived from the low-energy properties of strained HgTe, we can decrease the surface Fermi velocity v_F^{surf} by closing the bulk band gap $\mathcal{E}_{\text{gap}} \propto \delta m = |m - m_{c1}|$. However, this particular closing of the bulk band gap in the form of a “stamp” is not generic, although the surface bands become flat. Furthermore, the effect of the surface bands becoming flat is compensated by the screening properties of the polarization function which ultimately pre-empts the scenario of chiral symmetry breaking due to long-range Coulomb interactions, as discussed before. The role of local interactions in that model is more complicated, as discussed in section 5.5. The analysis of the susceptibility matrix $\chi(\mathbf{q})$ has shown that the leading instability of that model is surface ferromagnetism which opens a band gap for the surface Dirac fermions. However, since all eigenvalues of the susceptibility matrix are of the same order, this scenario only occurs when tuning the interaction U and the mass parameter m to a narrow parameter regime. In that sense, the gapped surface state in the minimal model is a fine-tuned situation and does not present a generic feature of the 2D surface states. Furthermore, as many other materials are described by a similar model Hamiltonian, the results presented here should be applicable to a range of topological insulator materials as well, but here again it comes down to numbers.

To conclude, we do not find a *generic* way to create flat surface bands which would create gapped surface states. Nevertheless, we have identified two *specific* mechanisms: First, a system may possess a chiral symmetry which gives rise to topologically protected bulk nodal lines and flat surface bands. Even if this chiral symmetry is only “approximate” in the sense that additional small terms break this symmetry, the surface bands are still reasonably flat so that one can hope to trigger a surface instability towards gapped surface states. The second mechanism is small inversion-symmetry breaking which creates a “stamp” of bulk valence and conduction bands. Here, the surface bands traversing the bulk band gap become increasingly flat as the stamp closes, but this happens on momentum scales that are usually hard to resolve in experiments due to the small inversion asymmetry.

APPENDIX A

Molecular Orbitals and Overlap Parameters

This appendix gives a brief derivation of the overlap parameters in terms of molecular orbitals. Since we are mainly interested in the electronic band structure of III-V or II-VI compounds such as HgTe, we restrict ourselves to the case of tetrahedrally bonded semiconductors. These semiconductors host valence electrons in s and p orbitals which are shown schematically in Fig. A.1. When two atoms are brought together the atomic orbitals overlap, and the tunneling between two atomic orbitals produces two hybridized orbitals. Because the two atoms are interchangeable, the symmetric superposition of the atomic orbitals forms the so-called **bonding orbital**, while the anti-symmetric superpositions forms the **anti-bonding orbital**. Note that there are two different ways for two p orbitals to overlap: p orbitals overlapping along the direction of the p orbitals form σ **bonds** [cf. Fig. A.1 (c)], whereas p orbitals overlapping in a direction perpendicular to the p orbitals form π **bonds** [cf. Fig. A.1 (d)].

Although the concept of bonding and anti-bonding orbitals was originally introduced for molecules, it can be extended to crystals by assuming that the orbitals of each atom in the crystal overlap only with the orbitals of its nearest neighbors. Furthermore, in a solid the bonding and anti-bonding orbitals are broadened into conduction and valence bands. For the case of tetrahedral semiconductors, the bonding orbitals are filled with electrons and become the valence bands, while the anti-bonding orbitals yield the conduction bands. The crystal structure, however, significantly affects the overlap between atomic orbitals.

The interaction between atomic orbitals lowers the energy of the bonding orbital by an amount V which is determined by the tight-binding Hamiltonian \mathcal{H}_0 between the different atomic orbitals of the atoms, whereas the energy of the anti-bonding orbital is increased by the same amount V . In both cases, the matrix element V of the tight-binding Hamiltonian is referred to as the **overlap parameter**. For a molecule containing only s and p valence electrons there are only four non-vanishing overlap parameters: $V_{ss\sigma}$, $V_{sp\sigma}$, $V_{pp\sigma}$, and $V_{pp\pi}$. To develop an explicit understanding of those overlap parameters, we consider a homopolar molecule, and we denote the atomic orbital on the first atom (designated as atom A in Fig. A.2) as $|\alpha\rangle$ and that on the other atom (B) as $|\beta\rangle$. These orbital states can be expressed in terms of radial wave functions and spherical harmonics $Y_{lm}(\theta, \phi)$, where the atom is located at the origin. Furthermore, we denote the direction vector from the first atom to the second atom as \mathbf{d} , and $\hat{\mathbf{d}} \equiv \mathbf{d}/|\mathbf{d}|$ is the unit vector along the direction of \mathbf{d} .

First, let us choose coordinates such that (i) the z axes are parallel to \mathbf{d} and (ii) the azimuthal angles ϕ are the same (see Fig. A.2). Within these coordinate systems, the spherical

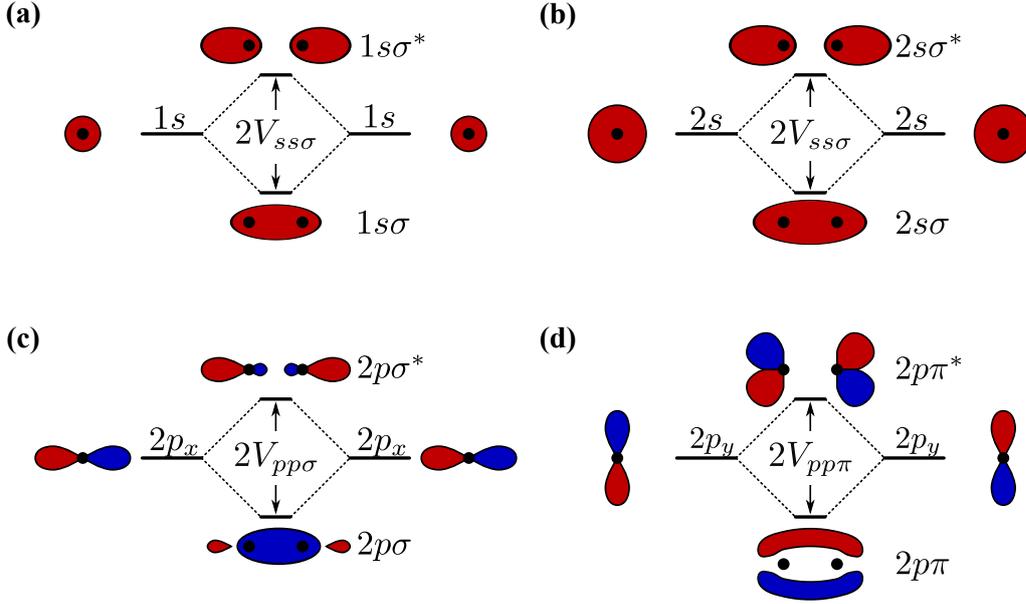


Figure A.1. Schematic picture of the overlap between s and p orbital wave functions of two atoms leading to molecular orbitals. **a, b)** Overlap of two $1s$ and $2s$ orbitals forming bonding and anti-bonding σ, σ^* orbitals. **c, d)** Overlap of two $2p_x$ and $2p_y$ orbitals along the x axis forming bonding and anti-bonding σ, σ^* and π, π^* orbitals, respectively.

harmonic wave function describing the two atomic orbitals $|\alpha\rangle$ and $|\beta\rangle$ are given by $Y_{lm}(\theta, \phi)$ and $Y_{l'm'}(\theta', \phi')$, respectively. Thus, the hopping Hamiltonian \mathcal{H}_0 has cylindrical symmetry with respect to \mathbf{d} , and therefore it cannot depend explicitly on ϕ . The corresponding matrix element is then proportional to the integral of the azimuthal wave function $\exp[i(m' - m)\phi]$ and vanishes unless $m = m'$:

$$\langle \alpha | \mathcal{H}_0 | \beta \rangle \propto \int_0^{2\pi} \frac{d\phi}{2\pi} e^{i(m' - m)\phi} = \delta_{m, m'}, \quad (\text{A.1})$$

where δ_{mn} denotes the Kronecker δ -symbol. For hopping in the z direction this **selection rule** leads to four non-vanishing and linearly independent overlap parameters between s and p electrons:

$$\begin{aligned} V_{ss\sigma} &\equiv \langle s | \mathcal{H}_0 | s \rangle, \\ V_{sp\sigma} &\equiv \langle s | \mathcal{H}_0 | p_z \rangle, \\ V_{pp\sigma} &\equiv \langle p_z | \mathcal{H}_0 | p_z \rangle, \\ V_{pp\pi} &\equiv \langle p_x | \mathcal{H}_0 | p_x \rangle. \end{aligned} \quad (\text{A.2})$$

Note that $\langle p_y | \mathcal{H}_0 | p_y \rangle = \langle p_x | \mathcal{H}_0 | p_x \rangle$ and that $\langle p_x | \mathcal{H}_0 | p_y \rangle = 0$ by symmetry. In general, those overlap parameters are labeled σ, π , and δ in analogy with the s ($l = 0$), p ($l = 1$), and d ($l = 2$) atomic wave functions.

In the above derivation of the overlap parameters, we have expanded the wave functions in terms of p orbitals along the Cartesian axes which are parallel or perpendicular to the bonding directions. However, in a tetrahedrally bonded semiconductor like HgTe or CdTe studied in the main text, the bond vectors \mathbf{d} linking nearest neighboring atoms are not identical to the Cartesian axes. Thus, a different choice for the coordinate systems is more convenient. To be precise, one typically chooses the crystallographic axes as the coordinate axes, and the spherical harmonics $Y_{lm}(\theta, \phi)$ of the atomic orbitals are then defined with respect to this fixed coordinate system. To construct the matrix elements of the tunneling Hamiltonian, we first decompose

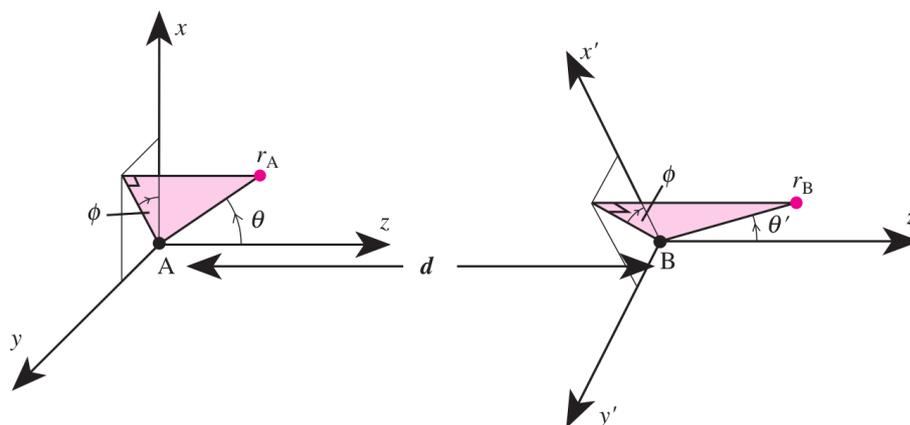


Figure A.2. Choice of the polar coordinate systems for atoms A and B in a two-atomic molecule. The z axes are chosen to be parallel to the vector \mathbf{d} connecting the two atoms, such that the azimuthal angle ϕ is the same for both coordinate systems. Figure after Ref. ⁷⁴.

the Cartesian p orbitals into the σ orbitals parallel and the π orbitals perpendicular to the bond vector \mathbf{d} , and afterwards we express the overlap between two p orbitals in terms of the overlap parameters $V_{ss\sigma}$, $V_{sp\sigma}$, $V_{pp\sigma}$, and $V_{pp\pi}$, as discussed below.

A.1 Overlap parameters between an s orbital and a p orbital

Let us first consider the matrix element $\langle s | \mathcal{H}_0 | p_j \rangle$ between the s orbital $|s\rangle$ and the $|p_j\rangle$ orbital, where $j = x, y, z$. Let $\hat{\mathbf{p}}$ be a unit vector along the direction of the p_j orbital. We start by decomposing the $|p_j\rangle$ orbital into two p orbitals parallel ($|p_\sigma\rangle$) and perpendicular ($|p_\pi\rangle$) to the direction vector $\hat{\mathbf{d}}$:

$$|p_j\rangle = (\hat{\mathbf{p}} \cdot \hat{\mathbf{d}}) |p_\sigma\rangle + (\hat{\mathbf{p}} \cdot \hat{\mathbf{n}}) |p_\pi\rangle, \quad (\text{A.3})$$

where $\hat{\mathbf{n}}$ is a unit vector normal to $\hat{\mathbf{d}}$ within the plane spanned by $\hat{\mathbf{d}}$ and $\hat{\mathbf{p}}$ [see Fig. A.3 (a)]. Using Eq. (A.2), we find that the Hamiltonian matrix element is then given by:

$$\langle s | \mathcal{H}_0 | p_j \rangle = (\hat{\mathbf{p}} \cdot \hat{\mathbf{d}}) \langle s | \mathcal{H}_0 | p_\sigma \rangle + (\hat{\mathbf{p}} \cdot \hat{\mathbf{n}}) \langle s | \mathcal{H}_0 | p_\pi \rangle = (\hat{\mathbf{p}} \cdot \hat{\mathbf{d}}) V_{sp\sigma}, \quad (\text{A.4})$$

where $V_{sp\sigma}$ is the overlap parameter describing the tunneling between an s orbital and a p orbital which is oriented parallel to \mathbf{d} . Note that the matrix element $\langle s | \mathcal{H}_0 | p_\pi \rangle$ vanishes, because the s orbital is invariant under rotations about the axis defined by the direction vector $\hat{\mathbf{d}}$, but the π orbital is not. The matrix elements between the p orbitals on atom A and the s orbital on atom B are obtained simply by inverting the direction of the bond vector \mathbf{d} , i.e., by changing the sign:

$$\langle p_j | \mathcal{H}_0 | s \rangle = -\langle s | \mathcal{H}_0 | p_j \rangle = (\hat{\mathbf{p}} \cdot \hat{\mathbf{d}}) V_{ps\sigma}, \quad (\text{A.5})$$

where, for brevity, we have defined $V_{ps\sigma} \equiv -V_{sp\sigma}$.

A.2 Overlap parameters between two p orbitals

The construction of the matrix element $\langle p_i | \mathcal{H}_0 | p_j \rangle$ between the p orbital $|p_i\rangle$ on the first atom and the p orbital $|p_j\rangle$ on the second requires a little bit more work. Consider two unit vectors $\hat{\mathbf{p}}_i$

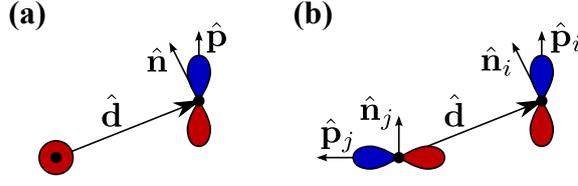


Figure A.3. To calculate the overlap between an s orbital and a p orbital state [panel (a)] or between two p orbital states [panel (b)] one has to decompose each p orbital into two p orbitals parallel and perpendicular to the unit vector $\hat{\mathbf{d}}$.

and $\hat{\mathbf{p}}_j$ along the directions of the two p orbitals, *i.e.*, along the x , y , or z axes [see Fig. A.3 (b)]. Both p orbitals can be decomposed into two σ and π orbitals parallel and perpendicular to $\hat{\mathbf{d}}$:

$$|p_i\rangle = (\hat{\mathbf{p}}_i \cdot \hat{\mathbf{d}})|p_{i\sigma}\rangle + (\hat{\mathbf{p}}_i \cdot \hat{\mathbf{n}}_i)|p_{i\pi}\rangle, \quad |p_j\rangle = (\hat{\mathbf{p}}_j \cdot \hat{\mathbf{d}})|p_{j\sigma}\rangle + (\hat{\mathbf{p}}_j \cdot \hat{\mathbf{n}}_j)|p_{j\pi}\rangle, \quad (\text{A.6})$$

where $\hat{\mathbf{n}}_i$ ($\hat{\mathbf{n}}_j$) denotes the unit vector perpendicular to $\hat{\mathbf{d}}$ in the plane spanned by $\hat{\mathbf{d}}$ and $\hat{\mathbf{p}}_i$ ($\hat{\mathbf{p}}_j$). The straightforward calculation of the matrix element between $|p_i\rangle$ and $|p_j\rangle$ then yields:

$$\begin{aligned} \langle p_i | \mathcal{H}_0 | p_j \rangle &= [(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{d}})\langle p_{i\sigma} | + (\hat{\mathbf{p}}_i \cdot \hat{\mathbf{n}}_i)\langle p_{i\pi} |] \mathcal{H}_0 [(\hat{\mathbf{p}}_j \cdot \hat{\mathbf{d}})|p_{j\sigma}\rangle + (\hat{\mathbf{p}}_j \cdot \hat{\mathbf{n}}_j)|p_{j\pi}\rangle] \\ &= (\hat{\mathbf{p}}_i \cdot \hat{\mathbf{d}})(\hat{\mathbf{p}}_j \cdot \hat{\mathbf{d}})\langle p_{i\sigma} | \mathcal{H}_0 | p_{j\sigma} \rangle + (\hat{\mathbf{p}}_i \cdot \hat{\mathbf{n}}_i)(\hat{\mathbf{p}}_j \cdot \hat{\mathbf{n}}_j)\langle p_{i\pi} | \mathcal{H}_0 | p_{j\pi} \rangle \\ &= (\hat{\mathbf{p}}_i \cdot \hat{\mathbf{d}})(\hat{\mathbf{p}}_j \cdot \hat{\mathbf{d}})V_{pp\sigma} + (\hat{\mathbf{p}}_i \cdot \hat{\mathbf{n}}_i)(\hat{\mathbf{p}}_j \cdot \hat{\mathbf{n}}_j)\delta_{ij}V_{pp\pi}, \end{aligned} \quad (\text{A.7})$$

where we have used the matrix elements $V_{pp\sigma}$ and $V_{pp\pi}$ describing the tunneling of electrons between p orbitals parallel and normal to the bond vector \mathbf{d} , respectively [cf. Eq. (A.2)]. Substituting the relation

$$(\hat{\mathbf{p}}_i \cdot \hat{\mathbf{n}}_i)(\hat{\mathbf{p}}_j \cdot \hat{\mathbf{n}}_j)\delta_{ij} = (\hat{\mathbf{p}}_i \cdot \hat{\mathbf{n}}_i)(\hat{\mathbf{p}}_j \cdot \hat{\mathbf{n}}_j)(\hat{\mathbf{n}}_i \cdot \hat{\mathbf{n}}_j) = [\hat{\mathbf{p}}_i - (\hat{\mathbf{p}}_i \cdot \hat{\mathbf{d}})\hat{\mathbf{d}}] \cdot [\hat{\mathbf{p}}_j - (\hat{\mathbf{p}}_j \cdot \hat{\mathbf{d}})\hat{\mathbf{d}}], \quad (\text{A.8})$$

we obtain:

$$\langle p_i | \mathcal{H}_0 | p_j \rangle = (\hat{\mathbf{p}}_i \cdot \hat{\mathbf{d}})(\hat{\mathbf{p}}_j \cdot \hat{\mathbf{d}})V_{pp\sigma} + [\hat{\mathbf{p}}_i - (\hat{\mathbf{p}}_i \cdot \hat{\mathbf{d}})\hat{\mathbf{d}}] \cdot [\hat{\mathbf{p}}_j - (\hat{\mathbf{p}}_j \cdot \hat{\mathbf{d}})\hat{\mathbf{d}}]V_{pp\pi} \quad (\text{A.9a})$$

$$= \begin{cases} (\hat{\mathbf{p}}_i \cdot \hat{\mathbf{d}})^2 V_{pp\sigma} + [1 - (\hat{\mathbf{p}}_i \cdot \hat{\mathbf{d}})^2] V_{pp\pi} & i = j \\ (\hat{\mathbf{p}}_i \cdot \hat{\mathbf{d}})(\hat{\mathbf{p}}_j \cdot \hat{\mathbf{d}})(V_{pp\sigma} - V_{pp\pi}) & i \neq j \end{cases} \quad (\text{A.9b})$$

A.3 Overlap parameters in diamond and zinc-blende lattices

Finally, let us summarize the above results and consider the overlap parameters $V_{ss\sigma}$, $V_{sp\sigma}$, $V_{ps\sigma}$, V_{xx} , and V_{xy} which are commonly used in the context of diamond or zinc-blende lattices. As discussed in the main text, in those crystal lattices the bond vectors between nearest neighboring atoms are given by

$$\begin{aligned} \mathbf{d}_1 &= \frac{a}{4}(1, 1, 1)^T, & \mathbf{d}_2 &= \frac{a}{4}(1, -1, -1)^T, \\ \mathbf{d}_3 &= \frac{a}{4}(-1, 1, -1)^T, & \mathbf{d}_4 &= \frac{a}{4}(-1, -1, 1)^T, \end{aligned} \quad (\text{A.10})$$

where a denotes the length of the crystallographic unit cell. Note that $|\mathbf{d}_i| = (a/4)\sqrt{3}$ for all $i = 1, \dots, 4$.

To compute the matrix elements of the tight-binding Hamiltonian \mathcal{H}_0 on the diamond or zinc-blende lattice, we assume that one of the two atoms, the anion denoted by $\nu = a$, is

located at the origin. Hence, the positions of the four nearest neighbors in those crystal lattices, the cations labeled by $\nu = c$, are then given by the vectors \mathbf{d}_j with $j = 1, \dots, 4$. First, let us consider the matrix element between the two s orbital states on the anion and cation, $\langle a, s, \sigma, \mathbf{k} | \mathcal{H}_0 | c, s, \sigma, \mathbf{k} \rangle$, where $\sigma \in \{\uparrow, \downarrow\}$ is the spin of the electron in the s orbitals, while \mathbf{k} denotes the crystal momentum in the first Brillouin zone. Since the overlap parameter depends on the angular momentum of the states involved and the distance between the atoms, we obtain:

$$\langle a, s, \sigma, \mathbf{k} | \mathcal{H}_0 | c, s, \sigma, \mathbf{k} \rangle = \sum_{j=1}^4 \langle s | \mathcal{H}_0 | s \rangle e^{i\mathbf{k} \cdot \mathbf{d}_j} = V_{ss} g_0(\mathbf{k}). \quad (\text{A.11})$$

In the last line, we have introduced the overlap parameter V_{ss} and the tight-binding function $g_0(\mathbf{k})$ as

$$V_{ss} \equiv \langle s | \mathcal{H}_0 | s \rangle = V_{ss\sigma} \quad (\text{A.12})$$

and

$$g_0(\mathbf{k}) \equiv \sum_{j=1}^4 e^{i\mathbf{k} \cdot \mathbf{d}_j}. \quad (\text{A.13})$$

Note that Eq. (A.11) is the same as Eq. (3.22) in the main text.

The matrix elements between s and p orbitals can be computed in a similar manner. For example, the overlap between the s orbital of the anion and the p_α orbital of the cation oriented in the e_α direction is given by:

$$\langle a, s, \sigma, \mathbf{k} | \mathcal{H}_0 | c, p_\alpha, \sigma, \mathbf{k} \rangle = \sum_{j=1}^4 \langle s | \mathcal{H}_0 | p_\alpha \rangle e^{i\mathbf{k} \cdot \mathbf{d}_j} = \sum_{j=1}^4 (\hat{\mathbf{p}}_\alpha \cdot \hat{\mathbf{d}}_j) V_{sp\sigma} e^{i\mathbf{k} \cdot \mathbf{d}_j}. \quad (\text{A.14})$$

Since $|\mathbf{d}_j| = (a/4)\sqrt{3}$ for all $j = 1, \dots, 4$, the above relation can be rewritten as

$$\langle a, s, \sigma, \mathbf{k} | \mathcal{H}_0 | c, p_\alpha, \sigma, \mathbf{k} \rangle = V_{sp} \sum_{j=1}^3 \text{sgn}(\hat{\mathbf{p}}_\alpha \cdot \hat{\mathbf{d}}) e^{i\mathbf{k} \cdot \mathbf{d}_j} = V_{sp} g_\alpha(\mathbf{k}), \quad (\text{A.15})$$

where we have introduced the overlap parameter

$$V_{sp} \equiv \frac{V_{sp\sigma}}{\sqrt{3}}, \quad (\text{A.16})$$

Furthermore, the sign which arises from the different orientations of the bond vectors \mathbf{d}_j defines three tight-binding functions

$$g_\alpha(\mathbf{k}) \equiv \frac{1}{4} \sum_{j=1}^4 \text{sgn}(\hat{\mathbf{p}}_\alpha \cdot \hat{\mathbf{d}}) e^{i\mathbf{k} \cdot \mathbf{d}_j}, \quad (\text{A.17})$$

where $\alpha \in \{x, y, z\} = \{1, 2, 3\}$ denotes the orientation of the p orbital state. Those functions are defined in Eq. (3.24) in the main text for $a = 1$.

Finally, the calculation of the overlap between different p orbitals of anion and cation can be performed in the same manner. First, the matrix element of p orbitals oriented in the same direction e_α is given by

$$\langle a, p_\alpha, \sigma, \mathbf{k} | \mathcal{H}_0 | c, p_\alpha, \sigma, \mathbf{k} \rangle = \sum_{j=1}^4 \langle p_\alpha | \mathcal{H}_0 | p_\alpha \rangle e^{i\mathbf{k} \cdot \mathbf{d}_j} = V_{xx} \sum_{j=1}^3 e^{i\mathbf{k} \cdot \mathbf{d}_j} = V_{xx} g_0(\mathbf{k}), \quad (\text{A.18})$$

where we have introduced the overlap parameter V_{xx} as

$$V_{xx} \equiv \frac{\langle p_\alpha | \mathcal{H}_0 | p_\alpha \rangle}{3} = \frac{V_{pp\sigma} + 2V_{pp\pi}}{3} \quad (\text{A.19})$$

using Eq. (A.9). Similarly, for the matrix elements of p orbitals oriented in different directions \mathbf{e}_α and \mathbf{e}_β we obtain:

$$\langle a, p_\alpha, \sigma, \mathbf{k} | \mathcal{H}_0 | c, p_\beta, \sigma, \mathbf{k} \rangle = \sum_{j=1}^4 \langle p_\alpha | \mathcal{H}_0 | p_\beta \rangle e^{i\mathbf{k} \cdot \mathbf{d}_j} = \sum_{j=1}^4 (\hat{\mathbf{p}}_\alpha \cdot \hat{\mathbf{d}}_j)(\hat{\mathbf{p}}_\beta \cdot \hat{\mathbf{d}}_j)(V_{pp\sigma} - V_{pp\pi}) e^{i\mathbf{k} \cdot \mathbf{d}_j}. \quad (\text{A.20})$$

This allows to define the following overlap parameter:

$$V_{xy} \equiv \frac{V_{pp\sigma} - V_{pp\pi}}{3}, \quad (\text{A.21})$$

so that the above matrix element can be formulated as

$$\langle a, p_\alpha, \sigma, \mathbf{k} | \mathcal{H}_0 | c, p_\beta, \sigma, \mathbf{k} \rangle = V_{xy} \frac{1}{4} \sum_{j=1}^4 \text{sgn}((\hat{\mathbf{p}}_\alpha \cdot \hat{\mathbf{d}}_j)(\hat{\mathbf{p}}_\beta \cdot \hat{\mathbf{d}}_j)) e^{i\mathbf{k} \cdot \mathbf{d}_j} = V_{xy} g_\gamma(\mathbf{k}), \quad (\text{A.22})$$

where γ is defined such that $|(\mathbf{e}_\alpha \times \mathbf{e}_\beta) \cdot \mathbf{e}_\gamma| = 1$. Note that the functions $g_\gamma(\mathbf{k})$ are the same functions as introduced before [cf. Eq. (A.17)].

APPENDIX B

Strain and Stress in Diamond and Zinc-Blende Crystals

In this appendix, we summarize the relations between strain and stress in diamond and zinc-blende-type crystal structures. To develop a better understanding, we first introduce the general concepts of strain and stress following the textbooks by Kittel⁷⁶ and Yu and Cardona⁷⁴. After that, we discuss the effect of tensile uniaxial stress along the high-symmetry [100], [110], and [111] directions of cubic crystals, before we continue with a discussion of tensile biaxial strain. We close with a brief summary of the consequences for the tight-binding calculations on the diamond and zinc-blende-type crystals like HgTe and CdTe discussed in the main text.

B.1 Elastic strain and stress

In general, the effect of a small crystal deformation on an atom, which is located at position \mathbf{r} , can be expressed in terms of a **displacement** vector $\mathbf{R}(\mathbf{r})$. If we consider three orthonormal vectors $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$, and $\hat{\mathbf{z}}$ with unit length in the unstrained solid and express the position of an individual atom in those coordinates as $\mathbf{r} = x\hat{\mathbf{x}} + y\hat{\mathbf{y}} + z\hat{\mathbf{z}}$, the displacement of that atom in the strained solid can be written as:

$$\mathbf{R}(\mathbf{r}) \equiv u_x(\mathbf{r})\hat{\mathbf{x}} + u_y(\mathbf{r})\hat{\mathbf{y}} + u_z(\mathbf{r})\hat{\mathbf{z}}. \quad (\text{B.1})$$

If the deformation is non-uniform, $u_x(\mathbf{r})$, $u_y(\mathbf{r})$, and $u_z(\mathbf{r})$ can be related to the local strains $\epsilon_{\alpha\beta}$ by a Taylor expansion of $\mathbf{R}(\mathbf{r})$ around $\mathbf{r} = 0$ using that $\mathbf{R}(0) = 0$. As a result, we may define the components of the so-called **strain tensor** ϵ by:

$$\epsilon_{\alpha\beta} \equiv \frac{1}{2} \left(\frac{\partial u_\alpha}{\partial x_\beta} + \frac{\partial u_\beta}{\partial x_\alpha} \right). \quad (\text{B.2})$$

The strain in the solid is generated by the applied **stress**, *i.e.*, by forces acting on a unit area in the solid. In general, there exist nine components $\sigma_{xx}, \sigma_{xy}, \sigma_{xz}, \sigma_{yx}, \sigma_{yy}, \sigma_{yz}, \sigma_{zx}, \sigma_{zy}, \sigma_{zz}$ of the **stress tensor** σ , where the first index represents the direction of the force, and the second subscript denotes the normal vector of the plane to which the force is applied. For example, σ_{xx} is a force applied in the x direction to a unit area in the yz plane, while σ_{xy} is a force applied in the x direction to a unit area in the xz plane. In a static situation, where the angular acceleration vanishes and where the total torque on the solid is zero, the number of independent stress

components reduces from nine to six components $\sigma_{xx}, \sigma_{yy}, \sigma_{zz}, \sigma_{xy}, \sigma_{yz}, \sigma_{zx}$, because $\sigma_{xy} = \sigma_{yx}, \sigma_{yz} = \sigma_{zy}$, and $\sigma_{zx} = \sigma_{xz}$.

The simplest relation between stress and strain is Hooke's law in which the applied force leads to a linear stretching of a spring. In a solid, there is a similar relation between stress and strain in terms of the **elastic stiffness constants** $C_{ij,kl}$ and the **elastic compliance constants** $S_{ij,kl}$:

$$\sigma_{ij} = \sum_{k,l} C_{ij,kl} \epsilon_{kl} \iff \epsilon_{kl} = \sum_{i,j} S_{ij,kl} \sigma_{ij}. \quad (\text{B.3})$$

In view of the large number of indices in the above relations, one usually simplifies the notation by introducing a new set of indices:

$$1 \equiv xx, \quad 2 \equiv yy, \quad 3 \equiv zz, \quad 4 \equiv yz, \quad 5 \equiv zx, \quad 6 \equiv xy. \quad (\text{B.4})$$

Consequently, the rank-two tensors σ and ϵ can be written as

$$\sigma = \begin{bmatrix} \sigma_{xx} & \sigma_{xy} & \sigma_{zx} \\ \sigma_{xy} & \sigma_{yy} & \sigma_{yz} \\ \sigma_{yz} & \sigma_{yz} & \sigma_{zz} \end{bmatrix} = \begin{bmatrix} \sigma_1 & \sigma_6 & \sigma_5 \\ \sigma_6 & \sigma_2 & \sigma_4 \\ \sigma_5 & \sigma_4 & \sigma_3 \end{bmatrix}, \quad (\text{B.5a})$$

and

$$\epsilon = \begin{bmatrix} \epsilon_{xx} & \epsilon_{xy} & \epsilon_{zx} \\ \epsilon_{xy} & \epsilon_{yy} & \epsilon_{yz} \\ \epsilon_{yz} & \epsilon_{yz} & \epsilon_{zz} \end{bmatrix} = \begin{bmatrix} \epsilon_1 & \epsilon_6/2 & \epsilon_5/2 \\ \epsilon_6/2 & \epsilon_2 & \epsilon_4/2 \\ \epsilon_5/2 & \epsilon_4/2 & \epsilon_3 \end{bmatrix}. \quad (\text{B.5b})$$

Note that by convention the off-diagonal elements of the strain tensor ϵ are defined with a factor of 2 in contrast to the stress tensor σ . A particular advantage of this contracted notation is that the fourth-rank stiffness tensor $C_{ij,kl}$ and accordingly the compliance tensor $S_{ij,kl}$ can be represented by a 6×6 matrix. Now let us define two six-component vectors ϵ and σ :

$$\epsilon \equiv (\epsilon_1, \epsilon_2, \epsilon_3, \epsilon_4, \epsilon_5, \epsilon_6)^T \equiv (\epsilon_{xx}, \epsilon_{yy}, \epsilon_{zz}, 2\epsilon_{yz}, 2\epsilon_{zx}, 2\epsilon_{xy})^T, \quad (\text{B.6a})$$

$$\sigma \equiv (\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5, \sigma_6)^T \equiv (\sigma_{xx}, \sigma_{yy}, \sigma_{zz}, \sigma_{yz}, \sigma_{zx}, \sigma_{xy})^T, \quad (\text{B.6b})$$

so that Eq. (B.3), which relates the strain and stress components, then takes the simple form of a matrix-vector product:

$$\sigma = C\epsilon \quad \text{and} \quad \epsilon = S\sigma. \quad (\text{B.7})$$

Hence, the compliance tensor $S = C^{-1}$ is just the inverse of the stiffness tensor C .

In general, there are 36 elastic stiffness constants C_{ij} and elastic compliance constants S_{ij} relating the six strain components ϵ_i to the stress components σ_j , but their number can be reduced by the following considerations. First, recall that the elastic energy density U is a quadratic function of the strain within the approximation of Hooke's law⁷⁶:

$$U = \frac{1}{2} \sum_{i,j=1}^6 u_{ij} \epsilon_i \epsilon_j, \quad (\text{B.8})$$

where the coefficients u_{ij} are related to the stiffness constants by virtue of Eq. (B.10). According to Hooke's law, the stress components σ_i are found from the derivative of the potential energy U with respect to the corresponding strain component:

$$\sigma_i \equiv \frac{\partial U}{\partial \epsilon_i} = u_{ii} \epsilon_i + \frac{1}{2} \sum_{j \neq i} (u_{ij} + u_{ji}) \epsilon_j. \quad (\text{B.9})$$

Note that only the symmetric combinations $(u_{ij} + u_{ji})/2$ enter the relations between strain and stress. Thus, the elastic stiffness constants are symmetric as well:

$$C_{ij} = \frac{u_{ij} + u_{ji}}{2} = C_{ji}. \quad (\text{B.10})$$

As a consequence, there are only 21 instead of 36 independent elastic stiffness constants.

The number of stiffness constants C_{ij} can be reduced even further if the solid possesses certain crystal symmetries. In view of the diamond and zinc-blende crystal structures discussed in the main text, we consider a cubic crystal which has at least four three-fold rotation axes in the $[111]$, $[1\bar{1}\bar{1}]$, $[\bar{1}1\bar{1}]$, and $[\bar{1}\bar{1}1]$ directions. The rotation of $2\pi/3$ about these four axes leads to an interchange of the x , y , and z axes as follows:

$$[111] : x \rightarrow z \rightarrow y \rightarrow x, \quad (\text{B.11a})$$

$$[1\bar{1}\bar{1}] : x \rightarrow -z \rightarrow -y \rightarrow x, \quad (\text{B.11b})$$

$$[\bar{1}1\bar{1}] : x \rightarrow z \rightarrow -y \rightarrow z \rightarrow x, \quad (\text{B.11c})$$

$$[\bar{1}\bar{1}1] : x \rightarrow -z \rightarrow y \rightarrow x. \quad (\text{B.11d})$$

Starting from the general expression (B.8) for the elastic energy, the first sequence for rotations about the body diagonal $[111]$ implies that $C_{11} = C_{22} = C_{33}$ as

$$C_{11}\epsilon_1^2 + C_{22}\epsilon_2^2 + C_{33}\epsilon_3^2 \rightarrow C_{11}\epsilon_2^2 + C_{22}\epsilon_3^2 + C_{33}\epsilon_1^2 \rightarrow \dots \quad (\text{B.12})$$

Similar relations between the different elastic stiffness constants are obtained from rotations about the other body diagonals. Taking all symmetries of the cubic crystal into account, the elastic energy has the following form:

$$U = \frac{C_{11}}{2}(\epsilon_1^2 + \epsilon_2^2 + \epsilon_3^2) + \frac{C_{44}}{2}(\epsilon_4^2 + \epsilon_5^2 + \epsilon_6^2) + \frac{C_{12}}{2}(\epsilon_1\epsilon_2 + \epsilon_2\epsilon_3 + \epsilon_3\epsilon_1). \quad (\text{B.13})$$

Thus, there are only three independent elastic stiffness constants C_{11} , C_{12} , and C_{44} in a cubic crystal, and the matrix of elastic stiffness constant takes the simple form

$$C = \begin{bmatrix} C_{11} & C_{12} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{11} & C_{12} & 0 & 0 & 0 \\ C_{12} & C_{12} & C_{11} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & C_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{44} \end{bmatrix}. \quad (\text{B.14})$$

Furthermore, by inverting the matrix of the elastic stiffness constants (B.14) we see that in cubic crystals the elastic stiffness constants C_{ij} and the elastic compliance constants S_{ij} are related by:

$$C_{11} - C_{12} = (S_{11} - S_{12})^{-1}, \quad (\text{B.15a})$$

$$C_{11} + 2C_{12} = (S_{11} + 2S_{12})^{-1}, \quad (\text{B.15b})$$

$$C_{44} = S_{44}^{-1}. \quad (\text{B.15c})$$

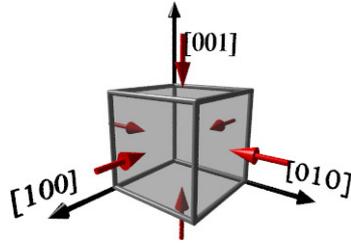


Figure B.1. Hydrostatic pressure is generated by the application of forces \mathbf{F} and $-\mathbf{F}$ with equal strength and opposite directions to the surfaces of an elementary cube within a solid.

B.2 Hydrostatic pressure

Let us consider a situation in which forces are applied along the x , y , and z axes to an elementary cube within a solid (see Fig. B.1). We assume that these forces come in pairs so that a force \mathbf{F} applied to one face will be balanced by another force $-\mathbf{F}$ applied to the opposing face, thus producing no acceleration of the solid. In addition, we require the off-diagonal forces to be equal so that there is no net torque. Consequently, the stress tensor σ is a symmetric second-rank tensor, whose elements are given by the forces per unit area applied to the elementary cube. Using the relation $\mathbf{F}_i = \sum_j \sigma_{ij} \mathbf{A}_j$ between the forces \mathbf{F}_i and the face normal vectors \mathbf{A}_j , we see that the stress tensor for **hydrostatic pressure** p is diagonal:

$$\sigma = \begin{bmatrix} -p & 0 & 0 \\ 0 & -p & 0 \\ 0 & 0 & -p \end{bmatrix}. \quad (\text{B.16})$$

Note that compressive stress is negative, while tensile stress is positive, according to our convention for the normal vectors \mathbf{A}_j pointing away from the solid (see Fig. B.1). In general, the application of hydrostatic pressure leads to a relative change in volume of the crystal proportional to p which is also known as **dilation** of the crystalline solid.

B.3 Uniaxial stress

If we consider a single pair of forces \mathbf{F} and $-\mathbf{F}$ with equal strength and opposite directions which are applied to surfaces of area A , whose normal vectors \mathbf{A} and $-\mathbf{A}$ are oriented along the same directions, the combined effect of the opposing forces \mathbf{F} and $-\mathbf{F}$ is to stretch the solid (see Fig. B.2). Since the forces are applied along the same axis, the solid is said to be under **tensile uniaxial stress**. In the following, we briefly discuss several realizations of such forms of stress.

B.3.1 Tensile uniaxial stress along the [100] direction

First, let us consider two forces $\mathbf{F} = F(1, 0, 0)^T$ and $-\mathbf{F}$ applied along the [100] direction to surfaces with normal vectors $\mathbf{A} = A(1, 0, 0)^T$ and $-\mathbf{A}$ [see Fig. B.2 (a)]. Using the relation $\mathbf{F}_i = \sum_j \sigma_{ij} \mathbf{A}_j$ for the second-rank stress tensor σ we find the following stress tensor σ :

$$\sigma = \begin{bmatrix} X & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (\text{B.17})$$

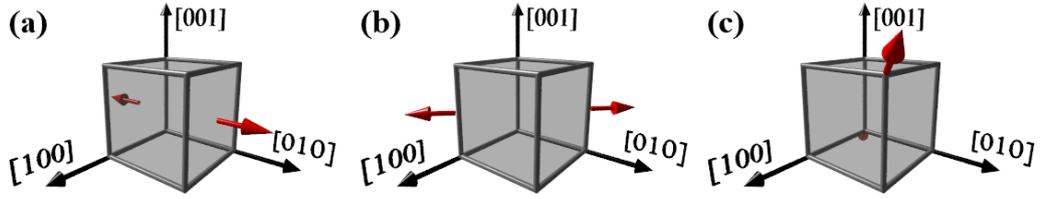


Figure B.2. The application of equal and opposite forces F and $-F$ (red arrows) to a solid along the (a) $[100]$ direction, (b) $[110]$ direction, and (c) $[111]$ direction leads to tensile uniaxial stress stretching the solid. The fractional change of volume of the crystal in (a) and (c) is equal to $(F/A)(S_{11} + 2S_{12})$, while in (b) the dilation is equal to $(F/A)(S_{11} + 3S_{12})$, where S_{ij} denote the elastic compliance constants for zinc-blende-type crystals.

Here, $X = F/A$ denotes the force per unit area. Note that according to our sign convention tensile stress is positive, *i.e.*, $X > 0$. The corresponding components of the strain tensor are obtained using Eq. (B.7):

$$\epsilon = \begin{bmatrix} XS_{11} & 0 & 0 \\ 0 & XS_{12} & 0 \\ 0 & 0 & XS_{12} \end{bmatrix}, \quad (\text{B.18})$$

where S_{11} and S_{12} denote the elastic compliance constants for a zinc-blende-type crystal. This matrix can be further decomposed into a sum of a diagonal matrix and a traceless shear strain tensor:

$$\epsilon = \frac{S_{11} + 2S_{12}}{3} \begin{bmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{bmatrix} + \frac{S_{11} - S_{12}}{3} \begin{bmatrix} 2X & 0 & 0 \\ 0 & -X & 0 \\ 0 & 0 & -X \end{bmatrix}. \quad (\text{B.19})$$

Physically, the diagonal matrix represents a hydrostatic pressure and thus leads to a dilation of the crystal by $X(S_{11} + 2S_{12})$ under tensile uniaxial stress along the $[001]$ direction, while the second shear strain tensor describes the relative deformation of the coordinate axes in the strained solid.

B.3.2 Tensile uniaxial stress along the $[110]$ direction

We encounter a different situation for a pair of forces $\mathbf{F} = (F/\sqrt{2})(1, 1, 0)^T$ and $-\mathbf{F}$ applied along the $[110]$ direction [see Fig. B.2 (b)]. Assuming that the surface normal vectors are equal to $\mathbf{A} = (A/\sqrt{2})(1, 1, 0)^T$ and $-\mathbf{A}$ we find the following stress tensor:

$$\sigma = \frac{X}{2} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (\text{B.20})$$

with $X = F/A$ the force per unit area. The corresponding strain tensor for uniaxial stress along the $[110]$ direction in a zinc-blende-type crystal is given by:

$$\epsilon = \frac{X}{2} \begin{bmatrix} S_{11} + S_{12} & S_{44}/2 & 0 \\ S_{44}/2 & S_{11} + S_{12} & 0 \\ 0 & 0 & 2S_{12} \end{bmatrix}. \quad (\text{B.21})$$

As before, the above strain tensor can also be decomposed into a hydrostatic pressure and a traceless shear strain tensor:

$$\epsilon = \frac{S_{11} + 3S_{12}}{2} \begin{bmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{bmatrix} + \frac{S_{11} - S_{12}}{2} \begin{bmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & -X \end{bmatrix} + \frac{S_{44}}{2} \begin{bmatrix} 0 & X & 0 \\ X & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (\text{B.22})$$

In contrast to tensile uniaxial stress along the [100] direction, we observe that the relative volume change for a uniaxial stress applied along the [110] direction is equal to $X(S_{11} + 3S_{12})$ rather than $X(S_{11} + 2S_{12})$. Furthermore, there is an additional deformation of the x and y axes in the crystal compared to the uniaxial strain along the [100] direction.

B.3.3 Tensile uniaxial stress along the [111] direction

Now, let us assume that a pair of forces $\mathbf{F} = (F/\sqrt{3})(1, 1, 1)^T$ and $-\mathbf{F}$ are applied along the [111] direction to surfaces of area A , whose normal vectors $\mathbf{A} = (A/\sqrt{3})(1, 1, 1)^T$ and $-\mathbf{A}$ are also oriented along the [111] directions [see Fig. B.2 (c)]. Here, we find that the stress tensor is given by

$$\sigma = \frac{X}{3} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad (\text{B.23})$$

and consequently the strain tensor corresponding to a uniaxial stress applied along the [111] direction of a zinc-blende-type crystal is given by

$$\epsilon = \frac{X}{3} \begin{bmatrix} S_{11} + 2S_{12} & S_{44}/2 & S_{44}/2 \\ S_{44}/2 & S_{11} + 2S_{12} & S_{44}/2 \\ S_{44}/2 & S_{44}/2 & S_{11} + 2S_{12} \end{bmatrix} \quad (\text{B.24})$$

with $X \equiv F/A$ the force per unit area. Here, ϵ can also be decomposed into the sum of a hydrostatic pressure and a traceless shear strain tensor as follows:

$$\epsilon = \frac{S_{11} + 2S_{12}}{3} \begin{bmatrix} X & 0 & 0 \\ 0 & X & 0 \\ 0 & 0 & X \end{bmatrix} + \frac{S_{44}}{2} \begin{bmatrix} 0 & X & X \\ X & 0 & X \\ X & X & 0 \end{bmatrix}. \quad (\text{B.25})$$

Here, the dilation of the crystal for a uniaxial stress along the [111] direction is also equal to $X(S_{11} + 2S_{12})$, just like for uniaxial stress along the [100] direction, but the coordinate axes in the strained solid are completely different from the other stress realizations.

B.4 Biaxial strain in the (001) plane

In quantum well heterostructures, there is usually a lattice mismatch between the different materials forming the quantum well layers which generically leads to strain in the sample. Let us assume that an epilayer of a zinc-blende-type crystal is grown epitaxially along the [001] direction on a substrate with similar crystal structure. The difference in the two lattice constants leads to a lattice mismatch

$$\Delta a \equiv \frac{a_s - a_e}{a_s}, \quad (\text{B.26})$$

where a_s and a_e denote the lattice constants of the substrate and the epilayer, respectively. This lattice mismatch between the substrate and the epilayer leads to either a tensile or a compressive strain on the epilayer. The stress corresponding to this strain is two-dimensional and lies within the (001) plane of the epilayer, while we assume that there is no stress on the epilayer along the [001] axis (see Fig. B.3). Consequently, such strain is called **biaxial strain** to distinguish it from uniaxial strain.

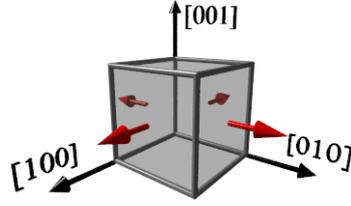


Figure B.3. Applying equal and opposite forces F and $-F$ along both the $[100]$ and $[010]$ directions leads to two-dimensional stress with the corresponding biaxial strain within the (001) epilayer. Although there is no stress in the $[001]$ direction, the application of biaxial stress leads to a hydrostatic strain on the solid.

The stress tensor σ describing biaxial stress in the (001) plane and the corresponding strain tensor ϵ obtained from Eq. (B.7) are given by

$$\sigma = X \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \epsilon = X \begin{bmatrix} S_{11} + S_{12} & 0 & 0 \\ 0 & S_{11} + S_{12} & 0 \\ 0 & 0 & 2S_{12} \end{bmatrix}, \quad (\text{B.27})$$

where the force per unit area, X , is generated by the difference in lattice constants, $\Delta a = (a_s - a_e)/a_s$. In terms of the elastic compliance constants S_{11} and S_{12} , this force on the solid due to lattice mismatch can be specified as

$$X = \frac{\Delta a}{S_{11} + S_{12}}. \quad (\text{B.28})$$

As a result, we obtain the following tensor describing biaxial strain:

$$\epsilon = \begin{bmatrix} \epsilon_1 & 0 & 0 \\ 0 & \epsilon_2 & 0 \\ 0 & 0 & \epsilon_3 \end{bmatrix} \quad \text{with} \quad \epsilon_1 = \epsilon_2 = \Delta a \quad \text{and} \quad \epsilon_3 = -\frac{2C_{12}}{C_{11}} \Delta a, \quad (\text{B.29})$$

where we have used that $(S_{11} + S_{12})^{-1} = -2C_{12}/C_{11}$ by virtue of Eq. (B.15). Note that we can also decompose the biaxial strain into the sum of a hydrostatic pressure and a uniaxial strain in the $[001]$ direction:

$$\epsilon = \frac{\Delta a}{a_s} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \frac{\Delta a}{a_s} \frac{C_{11} + 2C_{12}}{C_{11}} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (\text{B.30})$$

If the biaxial strain is compressive, $\Delta a/a_s < 0$, the hydrostatic pressure is also compressive, but the uniaxial strain in $[001]$ direction is tensile in nature.

B.5 Strained 3D HgTe samples

Finally, let us summarize the effects of biaxial strain on three-dimensional samples of HgTe. As discussed in the main text, if we grow HgTe epitaxially on a CdTe substrate, the lattice mismatch of the two materials induces **biaxial strain** within the epilayer of HgTe. Since the lattice constants of HgTe and CdTe are almost identical to each other,

$$a_{\text{HgTe}} = 6.46 \text{ \AA} \quad \text{and} \quad a_{\text{CdTe}} = 6.48 \text{ \AA}, \quad (\text{B.31})$$

the critical thickness for lattice relaxation of the HgTe epilayer is about 200 nm (Ref. 10). This implies that for thinner HgTe samples ($d_{\text{HgTe}} \ll 200$ nm) the epilayer adopts the in-plane lattice constant of the CdTe substrate. Although the lattice mismatch between CdTe and HgTe is rather small,

$$\Delta a = \frac{a_{\text{CdTe}} - a_{\text{HgTe}}}{a_{\text{CdTe}}} \approx 0.3\%, \quad (\text{B.32})$$

the resulting biaxial strain is sufficiently large to open a band gap between the Γ_{8v} and Γ_{8c} bands, as shown in the main text. In this appendix, we evaluate the strain tensor for biaxial strain and show how the bulk tight-binding calculations have to be modified when the solid is under the effect of strain.

By definition, the strain tensor ϵ leads to a displacement of an atom located at the position \mathbf{r} in the crystalline solid by a vector $\mathbf{R}(\mathbf{r}) = \epsilon \mathbf{r}$. Thus, for the biaxial strain in the (001) plane, as given by Eq. (B.29), we find that the changes $\delta \mathbf{d}_j$ of the nearest-neighbor vectors \mathbf{d}_j in the diamond or zinc-blende-type crystal are given by:

$$\delta \mathbf{d}_1 = \frac{a}{4} (\epsilon_1, \epsilon_2, \epsilon_3)^T = \frac{a}{4} \Delta a (1, 1, -2\alpha)^T, \quad (\text{B.33a})$$

$$\delta \mathbf{d}_2 = \frac{a}{4} (\epsilon_1, -\epsilon_2, -\epsilon_3)^T = \frac{a}{4} \Delta a (1, -1, 2\alpha)^T, \quad (\text{B.33b})$$

$$\delta \mathbf{d}_3 = \frac{a}{4} (-\epsilon_1, \epsilon_2, -\epsilon_3)^T = \frac{a}{4} \Delta a (-1, 1, 2\alpha)^T, \quad (\text{B.33c})$$

$$\delta \mathbf{d}_4 = \frac{a}{4} (-\epsilon_1, -\epsilon_2, \epsilon_3)^T = \frac{a}{4} \Delta a (-1, -1, 2\alpha)^T, \quad (\text{B.33d})$$

where, for brevity, we have defined $\alpha \equiv C_{12}/C_{11}$ as a ratio of elastic compliance constants. Thus, an atom located at \mathbf{d}_j in the unstrained solid will be displaced, and in the strained solid it is located at $\mathbf{d}'_1 = \mathbf{d}_1 + \Delta \mathbf{d}_1$:

$$\mathbf{d}'_1 = \frac{a}{4} (1, 1, 1)^T + \frac{a}{4} \Delta a (1, 1, -2\alpha)^T, \quad (\text{B.34a})$$

$$\mathbf{d}'_2 = \frac{a}{4} (1, -1, -1)^T + \frac{a}{4} \Delta a (1, -1, 2\alpha)^T, \quad (\text{B.34b})$$

$$\mathbf{d}'_3 = \frac{a}{4} (-1, 1, -1)^T + \frac{a}{4} \Delta a (-1, 1, 2\alpha)^T, \quad (\text{B.34c})$$

$$\mathbf{d}'_4 = \frac{a}{4} (-1, -1, 1)^T + \frac{a}{4} \Delta a (-1, -1, 2\alpha)^T. \quad (\text{B.34d})$$

Two effects of the applied strain are noteworthy: First, the overlap parameters in general depend on the atomic distance d as d^{-2} (cf. Ref. 74), so that the bare overlap parameters in the strained solid $V'_{ss\sigma}$, $V'_{sp\sigma}$, $V'_{pp\sigma}$, and $V'_{pp\pi}$ change in magnitude. Second, the angle between the bond vectors and the Cartesian axes changes in the strained solid, so that the definition of the overlap parameters V_{sp} , V_{xx} , and V_{xy} given in section A.3 is not valid anymore and the more general definitions of the matrix elements have to be used instead. For example, the matrix element describing the hopping between s and p orbitals introduced in the previous appendix has to be rewritten as:

$$\langle a, s, \sigma, \mathbf{k} | \mathcal{H}_0 | c, p_\alpha, \sigma, \mathbf{k} \rangle = \sum_{j=1}^4 (\hat{\mathbf{p}}_\alpha \cdot \hat{\mathbf{d}}'_j) V'_{sp\sigma} e^{i\mathbf{k} \cdot \mathbf{d}'_j}, \quad (\text{B.35})$$

where $V'_{sp\sigma}$ takes the change in the bond length into account, while $(\hat{\mathbf{p}}_\alpha \cdot \hat{\mathbf{d}}'_j)$ describes the modified bond angle with respect to the e_α direction. Similar relations hold for the other matrix elements as well. As a consequence, we have rewritten each matrix element of the 20×20 Bloch

Hamiltonian describing the hopping of electrons between anions and cations in the zinc-blende lattice of HgTe in terms of the bond vectors of the strained solid, and in the main text we have calculated numerically the electronic band structure and the corresponding density of states, taking into account both the changes in the bond lengths and the changes in the bond angles.

APPENDIX C

Topological Aspects of the Su-Schrieffer-Heeger Model

The **Su-Schrieffer-Heeger model** is a very successful theoretical model to study conjugated polymer chains^{183–185}. In their work, Su, Schrieffer, and Heeger focused on the coupling between the π electrons forming the valence band and the ionic motions along one-dimensional chains. This model exhibits topological excitations called **solitons** at domain walls coupling the two possible and equivalent configurations of bond-length alternation in the Peierls-distorted ground state. In this appendix, we briefly touch on the Su-Schrieffer-Heeger Hamiltonian and discuss its topological properties, which are quite similar to the boundary states of topological insulators discussed in the main text. In particular, we show that the different phases of that model can be classified by a winding number, *i.e.*, a topological invariant related to the first homotopy class $\pi_1(S^1) = \mathbb{Z}$. Furthermore, we give alternative formulations for the winding number in terms of the Bloch Hamiltonian which are used in chapter 5 to determine the topological aspects of the Kane-Mele and the Fu-Kane-Mele model, respectively.

C.1 The Su-Schrieffer-Heeger Hamiltonian

Historically, the Su-Schrieffer-Heeger model was developed for **polyacetylene**, $(\text{CH})_x$, which is a linear polymer consisting of weakly coupled chains of CH units forming a quasi-one-dimensional lattice^{183,184}. Three of the four carbon valence electrons are forming the three planar sp^2 -hybridized σ orbitals, where two of them are responsible for creating the one-dimensional chain between neighboring carbon atoms, while the third σ orbital forms a bond with the hydrogen atom. The remaining p_z orbitals of the carbon atom form the bonding and anti-bonding π , π^* orbitals. Since the optimal bond angle between the three σ orbitals is 120° , there exist two possible arrangements of the carbon atoms in nature, namely the ***cis*-(CH)_x** and the ***trans*-(CH)_x**, as shown in Fig. C.1¹⁸⁵. In both isomers, the σ bonds form completely filled bands, whereas the π orbitals lead to partially filled bands which are responsible for the electronic properties of the system.

Let us consider the ***trans*-(CH)_x** isomer. If all bonds were equally strong, this isomer would be a quasi-1D metal with a half-filled band. However, such a system is inherently unstable with respect to a **Peierls instability**, *i.e.*, a dimerization of the bonds in which adjacent CH groups move closer to each other and form single and double bonds. By symmetry, there are two such

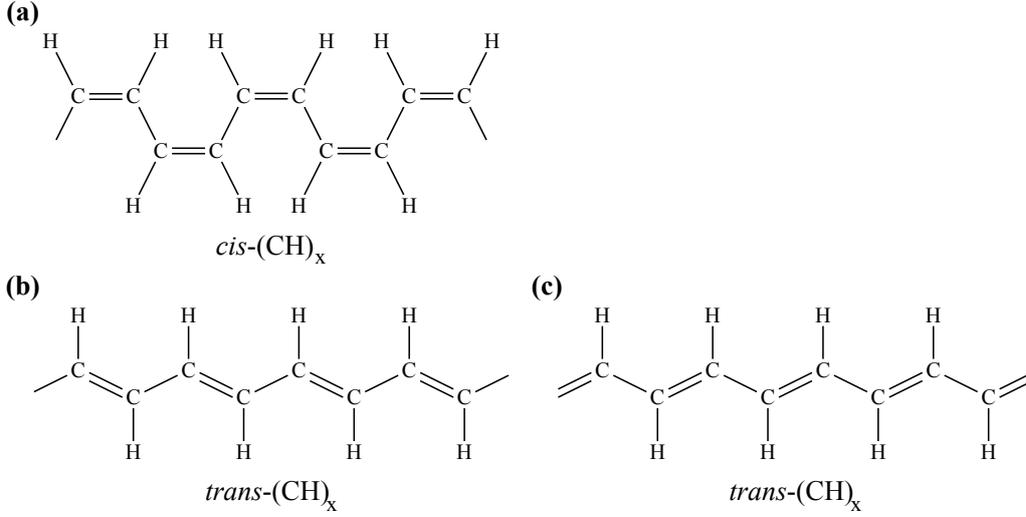


Figure C.1. Structure diagrams of the polyacetylene chains: Panel (a) shows the $cis-(CH)_x$ isomer, while panels (b, c) depict the two degenerate ground states of the $trans-(CH)_x$ isomer. Figure after Ref. 185.

states with lowest energy, as shown in Fig. C.1 (b, c). Note that the ground state of the $cis-(CH)_x$ isomer is non-degenerate, because there is no symmetry relating the two possible dimerizations. The two-fold degeneracy of the $trans-(CH)_x$ isomer leads to the existence of non-linear topological excitations or **solitons** at the domain walls between the two degenerate ground states^{183,184}. In the following, we consider an effective model for the $trans-(CH)_x$ isomer and discuss the properties of the Su-Schrieffer-Heeger model.

C.1.1 Tight-binding Hamiltonian for the Su-Schrieffer-Heeger model

In the first place, let us focus on a simple tight-binding Hamiltonian of the following form:

$$H = - \sum_j [t + (-1)^j \delta t] (c_j^\dagger c_{j+1} + \text{H.c.}), \quad (\text{C.1})$$

where t is the **overlap parameter** or **hopping amplitude**, and δt represents a modulation of the overlap parameter due to the dimerization of nearest neighbors. Furthermore, c_j and c_j^\dagger are the usual fermionic operators annihilating or creating a (spinless) electron at site j , respectively. Geometrically, the Peierls instability leads to a doubling of the crystallographic unit cell, thereby creating a sublattice structure of the model [see Fig. C.2 (a)]:

$$H = - \sum_j [(t + \delta t) a_j^\dagger b_j + (t - \delta t) (a_j^\dagger b_{j+1} + \text{H.c.})], \quad (\text{C.2})$$

where a, a^\dagger and b, b^\dagger are the annihilation and creation operators acting on the \mathfrak{A} and \mathfrak{B} sublattice, respectively. Let us introduce a two-component spinor $\psi_j = (a_j, b_j)^T$ with the Fourier decompositions

$$\psi_k = \frac{1}{\sqrt{N}} \sum_j e^{ikja} \psi_j, \quad \psi_j = \frac{1}{\sqrt{N}} \sum_k e^{-ikja} \psi_k, \quad (\text{C.3})$$

where a is the distance between adjacent atoms, and N denotes the number of two-atomic unit cells in the linear chain. It is straightforward to compute the corresponding Bloch Hamiltonian of the Su-Schrieffer-Heeger model:

$$\mathcal{H}(k) = [(t + \delta t) + (t - \delta t) \cos(ka)] \tau_x + (t - \delta t) \sin(ka) \tau_y = \mathbf{d}(k) \cdot \boldsymbol{\tau}, \quad (\text{C.4})$$

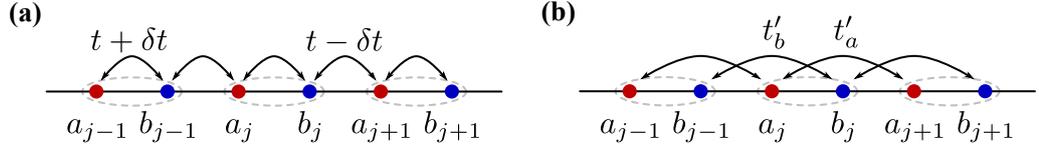


Figure C.2. a) Linear chain of polyacetylene with two-atomic sublattice indicated by red and blue filled circles. The Peierls instability leads to a dimerization of adjacent neighbors. b) Hopping of electrons between second-nearest neighbors with hopping amplitudes t'_a and t'_b , respectively.

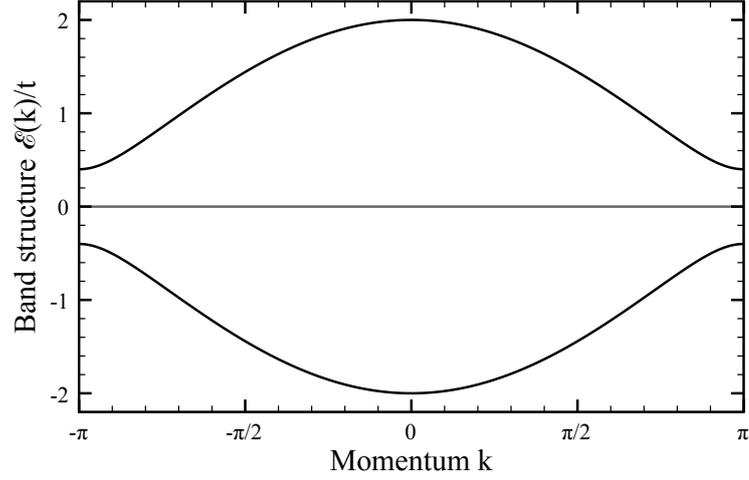


Figure C.3. Electronic band structure $\mathcal{E}_{\pm}(k)$ of the Su-Schrieffer-Heeger model for $t = 1$ and $|\delta t/t| = 0.2$. The bulk band gap of size $\mathcal{E}_{\text{gap}} = 2|\delta t/t|$ at $k = \pm\pi$ is clearly visible.

where, for convenience, we have introduced $\mathbf{d}(k) \in \mathbb{R}^3$ describing the 2×2 Bloch Hamiltonian $\mathcal{H}(k)$, and $\boldsymbol{\tau} = (\tau_x, \tau_y, \tau_z)^T$ is the vector of Pauli matrices acting on the sublattice degree of freedom [cf. Eq. (2.68)]. In that parametrization, the electronic band structure of that Bloch Hamiltonian is simply given by:

$$\mathcal{E}_{\pm}(k) = \pm|\mathbf{d}(k)| = \pm\sqrt{2}t\sqrt{(1 + \cos(ka)) + (\delta t/t)^2(1 - \cos(ka))}, \quad (\text{C.5})$$

where “ \pm ” denotes the conduction and valence band, respectively. Note that $\mathcal{E}_{\pm}(k)$ are independent of the sign of the hopping modulation δt which reflects the two-fold degeneracy of the ground state. Fig. C.3 shows the electronic band structure of the Su-Schrieffer-Heeger model with a clearly visible bulk band gap of $\mathcal{E}_{\text{gap}}/t = 2|\delta t/t|$.

C.1.2 Zero-energy bound states located at domains walls

For half-filling, we can expand the Bloch Hamiltonian (C.4) around the Fermi momentum $k_F = \pm\pi/a$ which leads to the following $(1 + 1)$ -dimensional Dirac Hamiltonian:

$$\mathcal{H}(k) = v_F k \tau_x + m \tau_z, \quad (\text{C.6})$$

where the Fermi velocity is given by $v_F \equiv (t + \delta t)a$, and the Dirac mass is equal to $m \equiv 2\delta t$. A particularly interesting aspect of that effective model is that it possesses a zero-energy bound state, as was discussed by Jackiw and Rebbi in 1976¹⁸⁶. Consider a spatially varying mass $m(x)$ such that $m(x \rightarrow \pm\infty) = \pm m_0$. By analyzing the above Dirac Hamiltonian in the continuum limit, *i.e.*, by replacing $k_x \rightarrow -i\hbar\partial_x$, Jackiw and Rebbi showed that this Dirac equation has

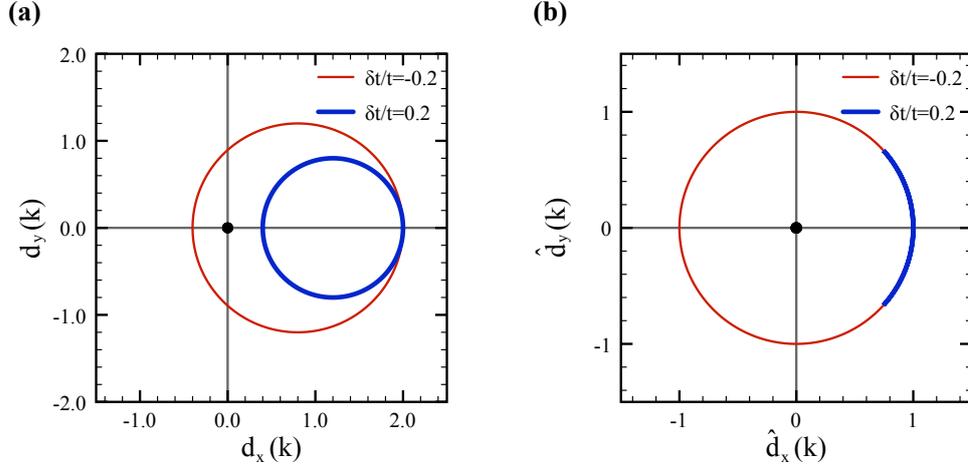


Figure C.4. **a)** Evolution of the vector $\mathbf{d}(k)$ as function of the crystal momentum k in the two-dimensional (d_x, d_y) plane. For $\delta t/t < 0$ the vector $\mathbf{d}(k)$ encircles the origin indicated by the thick black dot in the center, but not for $\delta t/t > 0$. **b)** Evolution of the normalized vector $\hat{\mathbf{d}}(k)$ as function of the momentum k . Here, the full revolution of the $\hat{\mathbf{d}}(k)$ vector for $\delta t/t < 0$ reflects the non-trivial winding of $\mathbf{d}(k)$, while for $\delta t/t > 0$ the winding is trivial as it only covers a part of the unit circle S^1 .

a unique and normalizable solution with zero energy which is exponentially localized at the domain wall, where $m(x)$ changes sign:

$$\psi_0(x) \propto \exp\left(-\int^x dx' \frac{m(x')}{\hbar v_F}\right) \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \quad (\text{C.7})$$

Note that for simplicity we have omitted the normalization constant. This bound state with zero energy is usually referred to as a **soliton** and can be interpreted as a topological excitation of the Su-Schrieffer-Heeger model^{183–186}. As shown by Jackiw and Rebbi, both the soliton and the anti-soliton as its partner have zero energy and thus form a degenerate doublet. Furthermore, they are characterized by a so-called **topological charge** which takes on the values $\pm\frac{1}{2}$ for the soliton and anti-soliton, respectively¹⁸⁶.

C.2 Topological origin of zero-energy bound states at domain walls

The existence of the zero-energy bound state at a domain wall can also be understood from the different topological properties of the model for $\delta t/t > 0$ and $\delta t/t < 0$, respectively. Fig. C.4 shows the evolution of the vector $\mathbf{d}(k)$ as function of crystal momentum k for positive and negative hopping modulation δt . Apparently, depending on the sign of δt , the vector $\mathbf{d}(k)$ may or may not enclose the origin, where $|\mathbf{d}(k)| = 0$ and the conduction and valence band touch. Note that the vector $\mathbf{d}(k)$ moves only within the two-dimensional (d_x, d_y) plane, because the model possesses a chiral symmetry $\hat{\Sigma} \equiv \tau_z$, *i.e.*, it anti-commutes with $\hat{\Sigma}$: $\{\mathcal{H}(k), \hat{\Sigma}\} = 0$. In the following, we explain the relation between the winding of the vector $\mathbf{d}(k)$ and the notion of topology in more detail, based on the work by Ryu and Hatsugai¹⁸.

To gain a better understanding of the topological aspects concerning the Su-Schrieffer-Heeger model, we consider a generic class of particle-hole symmetric two-band systems in one dimension whose Bloch Hamiltonian can conveniently be written as:

$$\mathcal{H}(k) = \mathbf{d}(k) \cdot \boldsymbol{\tau}, \quad \mathbf{d}(k) \in \mathbb{R}^3, \quad k \in [-\pi/a, \pi/a]. \quad (\text{C.8})$$

The corresponding eigenenergies of that Hamiltonian are simply given by $\mathcal{E}_\pm(k) = \pm|\mathbf{d}(k)|$, and the origin $\mathcal{O} \in \mathbb{R}^3$ corresponds to the gap-closing point, where the conduction and valence band touch and the bulk band gap vanishes. Note that for any momentum $k \in [-\pi/a, \pi/a]$, there exists a one-to-one correspondence between a point $\mathbf{d}(k)$ in the vector space \mathbb{R}^3 and the corresponding Bloch Hamiltonian $\mathcal{H}(k)$. Furthermore, since the boundaries of the Brillouin zone $\pm\pi/a$ are connected by the reciprocal lattice vector $2\pi/a$ and thus represent the same point, the one-dimensional Brillouin zone is topologically equivalent to the unit circle, *i.e.*, $[-\pi/a, \pi/a] \cong S^1$. We can therefore identify a **loop** $l : S^1 \rightarrow \mathbb{R}^3, k \mapsto \mathbf{d}(k)$ for each Bloch Hamiltonian $\mathcal{H}(k)$ which depends on the parameters of the system, and discuss the topological properties of the Su-Schrieffer-Heeger model in terms of those loops. The fundamental idea is now to consider a **homotopy**, *i.e.*, a continuous deformation of the Hamiltonian $\mathcal{H}(k)$ to a reference Hamiltonian which is known to exhibit zero-energy edge states¹⁸.

In the following, let us focus on a loop $l : S^1 \rightarrow \mathbb{R}^3, k \mapsto \mathbf{d}(k)$ which lies in a 2D plane that contains the origin $\mathcal{O} \in \mathbb{R}^3$. A necessary condition for this assumption is that the system possesses a **chiral symmetry** $\hat{\Sigma}$:

$$\{\hat{\Sigma}, \mathcal{H}(k)\} = 0 \quad \forall k \in S^1 \quad \iff \quad \mathcal{H}(k) \text{ respects chiral symmetry.} \quad (\text{C.9})$$

To prove that statement, we assume that a (normalized) vector $\hat{\mathbf{n}} \in \mathbb{R}^3$ exists such that

$$\mathbf{d}(k) \cdot \hat{\mathbf{n}} = 0 \quad \forall k \in S^1. \quad (\text{C.10})$$

Then we define a chiral symmetry operator $\hat{\Sigma} \equiv \hat{\mathbf{n}} \cdot \boldsymbol{\tau}$, so that $\mathcal{H}(k)$ anti-commutes with $\hat{\Sigma}$ for all momenta k in the Brillouin zone:

$$\{\hat{\Sigma}, \mathcal{H}(k)\} = \hat{n}_\alpha d_\beta(k) \{\tau_\alpha, \tau_\beta\} = \hat{n}_\alpha d_\beta(k) \times 2i\delta_{\alpha\beta}\tau_0 = 2i\hat{\mathbf{n}} \cdot \mathbf{d}(k) \tau_0 = 0, \quad (\text{C.11})$$

where the summation over indices α and β is implicit, and the last identity holds by Eq. (C.10). As a side remark, note that chiral symmetry also implies that the energy eigenvalues $\mathcal{E}_\pm(k)$ always appear in pairs with opposite sign since a global energy shift or bending of the bands due to a term $d_0(k)\tau_0$ is not allowed. Furthermore, note that we can restrict the loop l to lie within the (d_x, d_y) plane, because an arbitrary 2D plane can be rotated to the (d_x, d_y) plane by a unitary transformation. This stems from the fact that a global $SO(3)$ rotation in the vector space \mathbb{R}^3 as base manifold of $\mathbf{d}(k)$ amounts to a $SU(2)$ transformation of the Bloch Hamiltonian $\mathcal{H}(k)$.

Next, we further assume that we can continuously deform the loop $l : S^1 \rightarrow \mathbb{R}^2$, which from now on is taken to lie within the 2D (d_x, d_y) plane \mathbb{R}^2 , into a loop $l_c : S^1 \rightarrow S^1$ that lies on the unit circle centered around the origin \mathcal{O} :

$$l : S^1 \rightarrow \mathbb{R}^2, k \mapsto \mathbf{d}(k) \quad \xleftarrow{H} \quad l_c : S^1 \rightarrow S^1, k \mapsto \hat{\mathbf{d}}(k), \quad (\text{C.12})$$

where $H : S^1 \times [0, 1] \rightarrow \mathbb{R}^2$ is a **homotopy** on the product space of the unit circle S^1 with the unit interval $[0, 1]$ into the 2D space \mathbb{R}^2 such that $H(k, 0) = l(k)$ and $H(k, 1) = l_c(k)$. Most importantly, we assume that the loop l does not cross the origin \mathcal{O} during the deformation into l_c since $|\mathbf{d}(k = \mathcal{O})| = 0$ and thus $\hat{\mathbf{d}}(k = \mathcal{O})$ is not well-defined.

Instead of discussing the properties of the original Bloch Hamiltonian $\mathcal{H}(k)$ in terms of the vector $\mathbf{d}(k)$ we can now study loops $l_c : S^1 \rightarrow S^1$ on the unit circle. In particular, those loops can be classified by an integer **winding number** which counts the number of times the normalized vector $\hat{\mathbf{d}}(k)$ wraps around the unit circle as k goes through the Brillouin zone S^1 . This allows to define an equivalence relation between the loops and to classify them according to

their winding number. Mathematically, a more rigorous approach is to study the so-called **homotopy groups** of spheres, $\pi_n(S^m)$. In the simplest case $n = m = 1$, the homotopy group $\pi_1(S^1)$ also counts the number of times a mapping in the homotopy class wraps around the unit circle, just like in our case. Since the wrapping can occur in clockwise and counterclockwise directions, and wrapping in opposite directions cancel out, $\pi_1(S^1)$ is therefore an infinite cyclic group and is isomorphic to the group \mathbb{Z} of integers. The identification of the homotopy group $\pi_1(S^1)$ with the group of integers is often written as

$$\pi_1(S^1) = \mathbb{Z}. \quad (\text{C.13})$$

The non-triviality of this homotopy group is the mathematical reason for the existence of the zero-energy edge states at the domain walls of the Su-Schrieffer-Heeger model. In particular, the bulk-boundary correspondence implies that if two systems with different winding numbers ν_1 and ν_2 share a common interface, there exist $\Delta\nu = \nu_1 - \nu_2$ zero-energy bound states which are localized at the interface, as is further discussed below.

Geometrically, the two degenerate ground states of the Su-Schrieffer-Heeger model can be characterized by an integer number ν which can be identified as the **winding number** of the normalized vector $\hat{\mathbf{d}}(k)$. In other words, the integer ν counts how often the vector $\hat{\mathbf{d}}(k)$ wraps around the unit circle S^1 when going through the one-dimensional Brillouin zone $[-\pi/a, \pi/a]$. A general expression for this winding number is given by:

$$\nu = \frac{1}{2\pi} \int_{-\pi}^{\pi} dk \hat{\mathbf{n}} \cdot (\hat{\mathbf{d}}(k) \times \partial_k \hat{\mathbf{d}}(k)) = \frac{1}{2\pi} \int_{-\pi}^{\pi} dk \epsilon_{\alpha\beta\gamma} \hat{n}_\alpha \hat{d}_\beta(k) \partial_k \hat{d}_\gamma(k), \quad (\text{C.14})$$

where $\epsilon_{\alpha\beta\gamma}$ is the totally anti-symmetric tensor with $\epsilon_{123} = 1$, and the summation over α, β , and γ is implied. Note that the vector $\hat{\mathbf{d}}(k)$ does not cover the whole unit sphere, but only moves within a 2D plane. As discussed above, after performing a unitary transformation of the Bloch Hamiltonian $\mathcal{H}(k)$, the vector $\hat{\mathbf{d}}(k)$ lies in the (d_x, d_y) plane, and the winding number takes a simpler form:

$$\nu = \frac{1}{2\pi} \int_{-\pi}^{\pi} dk \epsilon_{\alpha\beta} \hat{d}_\alpha(k) \partial_k \hat{d}_\beta(k). \quad (\text{C.15})$$

Finally, for the Su-Schrieffer-Heeger model (C.4) one can compute the winding number ν analytically:

$$\nu = \frac{1 - \text{sgn}(\delta t/t)}{2} = \begin{cases} 0 & \text{for } \delta t/t > 0 \\ 1 & \text{for } \delta t/t < 0 \end{cases} \quad (\text{C.16})$$

This result is in agreement with the non-trivial winding number $\nu = 1$ of $\hat{\mathbf{d}}(k)$ for $\delta t/t < 0$ as shown in Fig. C.4. In particular, note that for $\delta t/t > 0$ the loop can be contracted onto a single point without crossing the origin in the 2D plane, and thus this loop is considered a trivial loop with $\nu = 0$.

Furthermore, the above result defines two classes of Hamiltonians with positive and negative hopping parameter δt which have two different winding numbers $\nu = 0$ and $\nu = 1$, respectively. To define those two classes more explicitly, we can define a “vacuum” Hamiltonian as $h_{\text{vac}}(k) \equiv h_{\text{vac}}$, where h_{vac} is an arbitrary Hermitian matrix which does not depend on the momentum k and which possesses chiral symmetry, *i.e.*, $\{\hat{\Sigma}, h_{\text{vac}}\} = 0$. Then h_{vac} describes a local system without any hopping between different sites. Such a trivial system with $\nu = 0$ can be used as a reference system, and we can classify all other systems with $\nu = 0$ as trivial as well, while those with $\nu = 1$ are non-trivial. Although this classification seems rather academic, such a topological classification has direct physical consequences. In particular, a domain wall between two topologically distinct phases exhibits zero-energy boundary states, such as the solitons and anti-solitons in the Su-Schrieffer-Heeger model.

C.2.1 Equivalent formulations of the winding number

A different, but equivalent formulation of the winding number ν can also be given in terms of the Bloch Hamiltonian $\mathcal{H}(k)$ as

$$\nu = \frac{1}{2} \int_{-\pi}^{\pi} \frac{dk}{2\pi i} \text{tr} \{ \hat{\Sigma} \mathcal{H}(k)^{-1} \partial_k \mathcal{H}(k) \}, \quad (\text{C.17})$$

where the additional prefactor $1/2$ accounts for the fact that we consider a two-band model, and $\hat{\Sigma}$ is the chiral symmetry operator introduced above. In that form, the winding number can be generalized to interacting systems as well, where one has to replace the Bloch Hamiltonian by the inverse of the fermion propagator, *i.e.*, $\mathcal{H}(k) \rightarrow G(k)^{-1}$ (Refs. 176,177).

Another equivalent formulation for the winding number of the Su-Schrieffer-Heeger model was given by Gurarie in Ref. 118 in terms of a complex-valued function $z(k)$. For the sake of concreteness, let us consider $\hat{n} = e_z$, so that the Bloch Hamiltonian takes the form $\mathcal{H}(k) = d_x(k)\tau_x + d_y(k)\tau_y$, and $z(k) = d_x(k) + id_y(k)$. As a result, we obtain:

$$\nu = \frac{1}{2} \int_{-\pi}^{\pi} \frac{dk}{2\pi i} \left(\frac{1}{z(k)} \frac{\partial z(k)}{\partial k} - \frac{1}{z(k)^*} \frac{\partial z(k)^*}{\partial k} \right). \quad (\text{C.18})$$

Here, note that $z(k)$ and $z(k)^*$ also appear as the off-diagonal matrix elements in the Bloch Hamiltonian $\mathcal{H}(k)$:

$$\mathcal{H}(k) = d_x(k)\tau_x + d_y(k)\tau_y = \frac{z(k)^*\tau_+ + z(k)\tau_-}{2} = \frac{1}{2} \begin{bmatrix} 0 & z(k)^* \\ z(k) & 0 \end{bmatrix}. \quad (\text{C.19})$$

C.3 Role of next-nearest neighbor hopping breaking chiral symmetry

In the context of topological insulators discussed in chapter 5, we have to discuss the effect of interactions between second-nearest neighbors which break the chiral symmetry of the Bloch Hamiltonian. In the Kane-Mele model on the graphene lattice or the Fu-Kane-Mele model on the diamond lattice, for example, such additional terms are given by the spin-independent hopping and the effective spin-orbit interaction between next-nearest neighbor atoms, or by the staggered sublattice potential. A common feature of those perturbations is that they couple to the τ_0 and/or τ_z matrices instead of τ_x and τ_y , and thereby remove the chiral symmetry of the model Hamiltonians.

To exemplify the role of such terms breaking chiral symmetry, let us consider the hopping of electrons between second-nearest neighbors with hopping amplitudes t'_a, t'_b as a small perturbation H' to the Su-Schrieffer-Heeger model [see Fig. C.2 (b)]. First, a straightforward Fourier decomposition of the operators leads to the new Bloch Hamiltonian $\mathcal{H}(k) + \mathcal{H}'(k)$, where $\mathcal{H}'(k)$ is given by

$$\mathcal{H}'(k) = (t'_a + t'_b) \cos(2ka)\tau_0 + (t'_a - t'_b) \cos(2ka)\tau_z \quad (\text{C.20})$$

with $\tau_0 = \mathbb{1}$ denoting the 2×2 identity matrix. The hopping between second-nearest neighbors actually breaks the chiral symmetry of the Su-Schrieffer-Heeger model, because for $\hat{\Sigma} = \tau_z$ we have:

$$\{\hat{\Sigma}, \tau_0\} = 2\tau_0 \quad \text{and} \quad \{\hat{\Sigma}, \tau_z\} = 2\tau_z. \quad (\text{C.21})$$

As a consequence, the loop l_c described by the unit vector $\hat{d}(k)$ of the Bloch Hamiltonian does not lie in a 2D plane, but rather lives on the unit sphere S^2 instead of the unit circle S^1 when k goes through the Brillouin zone. This modification has a profound consequence: The loop l_c

becomes contractible, because it is allowed to move on the whole sphere S^2 . Since the corresponding first homotopy class of S^2 is trivial,

$$\pi_1(S^2) = 0, \tag{C.22}$$

the Su-Schrieffer-Heeger model in the presence of next-nearest neighbor interactions does not show zero-energy bound states at domain walls. In that sense, the soliton and anti-soliton as topological excitations are protected by chiral symmetry, and generically any symmetry-breaking perturbation such as long-range hopping processes destroy the zero-energy bound states. Furthermore, this also shows that, in general, there are no zero-energy edge states present on the surfaces of the Kane-Mele model or the Fu-Kane-Mele model, with the possible exception of the time-reversal invariant momenta in the particle-hole symmetric case. In the main text, this was also confirmed by explicit tight-binding calculations of those models.

Bibliography

- [1] K. von Klitzing, G. Dorda, and M. Pepper. “New Method for High-Accuracy Determination of the Fine-Structure Constant Based on Quantized Hall Resistance.” *Phys. Rev. Lett.* **45**, 494 (1980). doi: 10.1103/PhysRevLett.45.494.
- [2] F. D. M. Haldane. “Model for a Quantum Hall Effect without Landau Levels: Condensed-Matter Realization of the ‘Parity Anomaly’.” *Phys. Rev. Lett.* **61**, 2015 (1988). doi: 10.1103/PhysRevLett.61.2015.
- [3] C. L. Kane and E. J. Mele. “Quantum Spin Hall Effect in Graphene.” *Phys. Rev. Lett.* **95**, 226801 (2005). doi: 10.1103/PhysRevLett.95.226801.
- [4] C. L. Kane and E. J. Mele. “ Z_2 Topological Order and the Quantum Spin Hall Effect.” *Phys. Rev. Lett.* **95**, 146802 (2005). doi: 10.1103/PhysRevLett.95.146802.
- [5] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang. “Quantum Spin Hall Effect and Topological Phase Transition in HgTe Quantum Wells.” *Science* **314**, 1757 (2006). doi: 10.1126/science.1133734.
- [6] M. König, *et al.* “Quantum Spin Hall Insulator State in HgTe Quantum Wells.” *Science* **318**, 766 (2007). doi: 10.1126/science.1148047.
- [7] L. Fu, C. L. Kane, and E. J. Mele. “Topological Insulators in Three Dimensions.” *Phys. Rev. Lett.* **98**, 106803 (2007). doi: 10.1103/PhysRevLett.98.106803.
- [8] L. Fu and C. L. Kane. “Topological insulators with inversion symmetry.” *Phys. Rev. B* **76**, 045302 (2007). doi: 10.1103/PhysRevB.76.045302.
- [9] X. Dai, *et al.* “Helical edge and surface states in HgTe quantum wells and bulk insulators.” *Phys. Rev. B* **77**, 125319 (2008). doi: 10.1103/PhysRevB.77.125319.
- [10] C. Brüne, *et al.* “Quantum Hall Effect from the Topological Surface States of Strained Bulk HgTe.” *Phys. Rev. Lett.* **106**, 126803 (2011). doi: 10.1103/PhysRevLett.106.126803.
- [11] X.-L. Qi, T. L. Hughes, and S.-C. Zhang. “Topological field theory of time-reversal invariant insulators.” *Phys. Rev. B* **78**, 195424 (2008). doi: 10.1103/PhysRevB.78.195424.
- [12] A. M. Essin, J. E. Moore, and D. Vanderbilt. “Magnetoelectric Polarizability and Axion Electrodynamics in Crystalline Insulators.” *Phys. Rev. Lett.* **102**, 146805 (2009). doi: 10.1103/PhysRevLett.102.146805.
- [13] X.-L. Qi and S.-C. Zhang. “Topological insulators and superconductors.” *Rev. Mod. Phys.* **83**, 1057 (2011). doi: 10.1103/RevModPhys.83.1057.

-
- [14] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs. “Quantized Hall Conductance in a Two-Dimensional Periodic Potential.” *Phys. Rev. Lett.* **49**, 405 (1982). doi: 10.1103/PhysRevLett.49.405.
- [15] D. V. Khveshchenko and H. Leal. “Excitonic instability in layered degenerate semimetals.” *Nucl. Phys. B* **687**, 323 (2004). doi: 10.1016/j.nuclphysb.2004.03.020.
- [16] D. T. Son. “Quantum critical point in graphene approached in the limit of infinitely strong Coulomb interaction.” *Phys. Rev. B* **75**, 235423 (2007). doi: 10.1103/PhysRevB.75.235423.
- [17] J. E. Drut and T. A. Lähde. “Is Graphene in Vacuum an Insulator?” *Phys. Rev. Lett.* **102**, 026802 (2009). doi: 10.1103/PhysRevLett.102.026802.
- [18] S. Ryu and Y. Hatsugai. “Topological Origin of Zero-Energy Edge States in Particle-Hole Symmetric Systems.” *Phys. Rev. Lett.* **89**, 077002 (2002). doi: 10.1103/PhysRevLett.89.077002.
- [19] M. Sitte, A. Rosch, E. Altman, and L. Fritz. “Topological Insulators in Magnetic Fields: Quantum Hall Effect and Edge Channels with a Nonquantized θ Term.” *Phys. Rev. Lett.* **108**, 126807 (2012). doi: 10.1103/PhysRevLett.108.126807.
- [20] M. Sitte, A. Rosch, and L. Fritz. “Tuning the Fermi velocity of Dirac cones: Towards an anomalous quantum Hall effect on the surfaces of topological insulators?” In preparation for publication (2012).
- [21] Nobelprize.org. “The Nobel Prize in Physics 1985.” Published online (May 2012).
- [22] R. B. Laughlin. “Quantized Hall conductivity in two dimensions.” *Phys. Rev. B* **23**, 5632 (1981). doi: 10.1103/PhysRevB.23.5632.
- [23] A. Tzalenchuk, *et al.* “Towards a quantum resistance standard based on epitaxial graphene.” *Nat. Nano.* **5**, 186 (2010). doi: 10.1038/nnano.2009.474.
- [24] H. L. Störmer. “Two-dimensional electron correlation in high magnetic fields.” *Phys. B* **177**, 401 (1992). doi: 10.1016/0921-4526(92)90138-I.
- [25] R. Willett, *et al.* “Observation of an even-denominator quantum number in the fractional quantum Hall effect.” *Phys. Rev. Lett.* **59**, 1776 (1987). doi: 10.1103/PhysRevLett.59.1776.
- [26] Nobelprize.org. “The Nobel Prize in Physics 1998.” Published online (May 2012).
- [27] M. König, *et al.* “The Quantum Spin Hall Effect: Theory and Experiment.” *J. Phys. Soc. Jpn.* **77**, 031007 (2008). doi: 10.1143/JPSJ.77.031007.
- [28] M. Z. Hasan and C. L. Kane. “Colloquium: Topological insulators.” *Rev. Mod. Phys.* **82**, 3045 (2010). doi: 10.1103/RevModPhys.82.3045.
- [29] Wikimedia Commons. “Quantum Hall Effect.” Published online (May 2012).
- [30] S. M. Girvin. “The Quantum Hall Effect: Novel Excitations and Broken Symmetries.” arXiv:cond-mat/9907002 (1999).
- [31] M. O. Goerbig. “Quantum Hall Effects.” arXiv:0909.1998 (2009).

- [32] G. Watson. “Hall conductance as a topological invariant.” *Cont. Phys.* **37**, 127 (1996). doi: 10.1080/00107519608230340.
- [33] G. D. Mahan. *Many-Particle Physics*. Third edition (Kluwer Academic Publishers, 2000). ISBN 0-306-46338-5.
- [34] Q. Niu, D. J. Thouless, and Y.-S. Wu. “Quantized Hall conductance as a topological invariant.” *Phys. Rev. B* **31**, 3372 (1985). doi: 10.1103/PhysRevB.31.3372.
- [35] J. E. Avron, R. Seiler, and B. Simon. “Homotopy and Quantization in Condensed Matter Physics.” *Phys. Rev. Lett.* **51**, 51 (1983). doi: 10.1103/PhysRevLett.51.51.
- [36] Y. Hatsugai. “Chern number and edge states in the integer quantum Hall effect.” *Phys. Rev. Lett.* **71**, 3697 (1993). doi: 10.1103/PhysRevLett.71.3697.
- [37] X.-L. Qi and S.-C. Zhang. “The quantum spin Hall effect and topological insulators.” *Phys. Today* **63**, 33 (2010). doi: 10.1063/1.3293411.
- [38] L. Fu and C. L. Kane. “Time reversal polarization and a \mathbb{Z}_2 adiabatic spin pump.” *Phys. Rev. B* **74**, 195312 (2006). doi: 10.1103/PhysRevB.74.195312.
- [39] T. Fukui and Y. Hatsugai. “Quantum Spin Hall Effect in Three Dimensional Materials: Lattice Computation of \mathbb{Z}_2 Topological Invariants and Its Application to Bi and Sb.” *J. Phys. Soc. Jpn.* **76**, 053702 (2007). doi: 10.1143/JPSJ.76.053702.
- [40] J. E. Moore and L. Balents. “Topological invariants of time-reversal-invariant band structures.” *Phys. Rev. B* **75**, 121306 (2007). doi: 10.1103/PhysRevB.75.121306.
- [41] T. Fukui, T. Fujiwara, and Y. Hatsugai. “Topological Meaning of \mathbb{Z}_2 Numbers in Time Reversal Invariant Systems.” *J. Phys. Soc. Jpn* **77**, 123705 (2008). doi: 10.1143/JPSJ.77.123705.
- [42] R. Roy. “ \mathbb{Z}_2 classification of quantum spin Hall systems: An approach using time-reversal invariance.” *Phys. Rev. B* **79**, 195321 (2009). doi: 10.1103/PhysRevB.79.195321.
- [43] R. Roy. “Topological phases and the quantum spin Hall effect in three dimensions.” *Phys. Rev. B* **79**, 195322 (2009). doi: 10.1103/PhysRevB.79.195322.
- [44] R. Roy. “Characterization of three-dimensional topological insulators by two-dimensional invariants.” *New J. Phys.* **12**, 065009 (2010). doi: 10.1088/1367-2630/12/6/065009.
- [45] Z. Wang, X.-L. Qi, and S.-C. Zhang. “Equivalent topological invariants of topological insulators.” *New J. Phys.* **12**, 065007 (2010). doi: 10.1088/1367-2630/12/6/065007.
- [46] A. A. Soluyanov and D. Vanderbilt. “Computing topological invariants without inversion symmetry.” *Phys. Rev. B* **83**, 235401 (2011). doi: 10.1103/PhysRevB.83.235401.
- [47] D. N. Sheng, Z. Y. Weng, L. Sheng, and F. D. M. Haldane. “Quantum Spin-Hall Effect and Topologically Invariant Chern Numbers.” *Phys. Rev. Lett.* **97**, 036808 (2006). doi: 10.1103/PhysRevLett.97.036808.
- [48] C. Wu, B. A. Bernevig, and S.-C. Zhang. “Helical Liquid and the Edge of Quantum Spin Hall Systems.” *Phys. Rev. Lett.* **96**, 106401 (2006). doi: 10.1103/PhysRevLett.96.106401.

- [49] C. Xu and J. E. Moore. “Stability of the quantum spin Hall effect: Effects of interactions, disorder, and \mathbb{Z}_2 topology.” *Phys. Rev. B* **73**, 045322 (2006). doi: 10.1103/PhysRevB.73.045322.
- [50] H. B. Nielsen and M. Ninomiya. “The Adler-Bell-Jackiw anomaly and Weyl fermions in a crystal.” *Phys. Lett. B* **130**, 389 (1983). doi: 10.1016/0370-2693(83)91529-0.
- [51] K. Nomura, M. Koshino, and S. Ryu. “Topological Delocalization of Two-Dimensional Massless Dirac Fermions.” *Phys. Rev. Lett.* **99**, 146806 (2007). doi: 10.1103/PhysRevLett.99.146806.
- [52] Y. L. Chen, *et al.* “Experimental Realization of a Three-Dimensional Topological Insulator, Bi_2Te_3 .” *Science* **325**, 178 (2009). doi: 10.1126/science.1173034.
- [53] Y. Xia, *et al.* “Observation of a large-gap topological-insulator class with a single Dirac cone on the surface.” *Nat. Phys.* **5**, 398 (2009). doi: 10.1038/nphys1274.
- [54] 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.” *Nat. Phys.* **5**, 438 (2009). doi: 10.1038/nphys1270.
- [55] Y. L. Chen, *et al.* “Massive Dirac Fermion on the Surface of a Magnetically Doped Topological Insulator.” *Science* **329**, 659 (2010). doi: 10.1126/science.1189924.
- [56] H.-M. Guo and M. Franz. “Three-Dimensional Topological Insulators on the Pyrochlore Lattice.” *Phys. Rev. Lett.* **103**, 206805 (2009). doi: 10.1103/PhysRevLett.103.206805.
- [57] W. Witczak-Krempa and Y. B. Kim. “Topological and magnetic phases of interacting electrons in the pyrochlore iridates.” *Phys. Rev. B* **85**, 045124 (2012). doi: 10.1103/PhysRevB.85.045124.
- [58] X. Wan, A. M. Turner, A. Vishwanath, and S. Y. Savrasov. “Topological semimetal and Fermi-arc surface states in the electronic structure of pyrochlore iridates.” *Phys. Rev. B* **83**, 205101 (2011). doi: 10.1103/PhysRevB.83.205101.
- [59] M. Kargarian, J. Wen, and G. A. Fiete. “Competing exotic topological insulator phases in transition-metal oxides on the pyrochlore lattice with distortion.” *Phys. Rev. B* **83**, 165112 (2011). doi: 10.1103/PhysRevB.83.165112.
- [60] S. Chadov, *et al.* “Tunable multifunctional topological insulators in ternary Heusler compounds.” *Nat. Mat.* **9**, 541 (2010). doi: 10.1038/nmat2770.
- [61] H. Lin, *et al.* “Half-Heusler ternary compounds as new multifunctional experimental platforms for topological quantum phenomena.” *Nat. Mater.* **9**, 546 (2010). doi: 10.1038/nmat2771.
- [62] M. R. Zirnbauer. “Riemannian symmetric superspaces and their origin in random-matrix theory.” *Journal of Mathematical Physics* **37**, 4986 (1996). doi: 10.1063/1.531675.
- [63] A. Altland and M. R. Zirnbauer. “Nonstandard symmetry classes in mesoscopic normal-superconducting hybrid structures.” *Phys. Rev. B* **55**, 1142 (1997). doi: 10.1103/PhysRevB.55.1142.
- [64] P. Heinzner, A. Huckleberry, and M. R. Zirnbauer. “Symmetry Classes of Disordered Fermions.” *Commun. Math. Phys.* **257**, 725 (2005). doi: 10.1007/s00220-005-1330-9.

- [65] A. P. Schnyder, S. Ryu, A. Furusaki, and A. W. W. Ludwig. “Classification of Topological Insulators and Superconductors.” *AIP Conf. Proc.* **1134**, 10 (2008). doi: 10.1063/1.3149481.
- [66] A. P. Schnyder, S. Ryu, A. Furusaki, and A. W. W. Ludwig. “Classification of topological insulators and superconductors in three spatial dimensions.” *Phys. Rev. B* **78**, 195125 (2008). doi: 10.1103/PhysRevB.78.195125.
- [67] S. Ryu, A. P. Schnyder, A. Furusaki, and A. W. W. Ludwig. “Topological insulators and superconductors: tenfold way and dimensional hierarchy.” *New J. Phys.* **12**, 065010 (2010). doi: 10.1088/1367-2630/12/6/065010.
- [68] C. Day. “Quantum Spin Hall Effect Shows up in a Quantum Well Insulator, Just as Predicted.” *Phys. Today* **61**, 19 (2008). doi: 10.1063/1.2835139.
- [69] D. Hsieh, *et al.* “A topological Dirac insulator in a quantum spin Hall phase.” *Nature* **452**, 970 (2008). doi: 10.1038/nature06843.
- [70] J. C. Y. Teo, L. Fu, and C. L. Kane. “Surface states and topological invariants in three-dimensional topological insulators: Application to $\text{Bi}_{1-x}\text{Sb}_x$.” *Phys. Rev. B* **78**, 045426 (2008). doi: 10.1103/PhysRevB.78.045426.
- [71] D. Hsieh, *et al.* “Observation of Unconventional Quantum Spin Textures in Topological Insulators.” *Science* **323**, 919 (2009). doi: 10.1126/science.1167733.
- [72] S.-Y. Xu, *et al.* “Topological Phase Transition and Texture Inversion in a Tunable Topological Insulator.” *Science* **332**, 560 (2011). doi: 10.1126/science.1201607.
- [73] F. Viot, R. Hayn, M. Richter, and J. van den Brink. “Metacinnabar (β -HgS): A Strong 3D Topological Insulator with Highly Anisotropic Surface States.” *Phys. Rev. Lett.* **106**, 236806 (2011). doi: 10.1103/PhysRevLett.106.236806.
- [74] P. Y. Yu and M. Cardona. *Fundamentals of Semiconductors*. Graduate Texts in Physics, fourth edition (Springer-Verlag, 2010). ISBN 978-3-642-00709-5.
- [75] R. Winkler. *Spin-orbit Coupling Effects in Two-Dimensional Electron and Hole Systems*, volume 191 of *Springer Tracts in Modern Physics* (Springer-Verlag, 2003). ISBN 978-3-540-01187-3.
- [76] C. Kittel. *Introduction to Solid State Physics*. Eighth edition (Wiley, 2004). ISBN 978-0-471-41526-8.
- [77] D. J. Chadi and M. L. Cohen. “Tight-binding calculations of the valence bands of diamond and zincblende crystals.” *phys. stat. sol. (b)* **68**, 405 (1975). doi: 10.1002/pssb.2220680140.
- [78] D. J. Chadi. “Spin-orbit splitting in crystalline and compositionally disordered semiconductors.” *Phys. Rev. B* **16**, 790 (1977). doi: 10.1103/PhysRevB.16.790.
- [79] A. Kobayashi, O. F. Sankey, and J. D. Dow. “Chemical trends for defect energy levels in $\text{Hg}_{(1-x)}\text{Cd}_x\text{Te}$.” *Phys. Rev. B* **25**, 6367 (1982). doi: 10.1103/PhysRevB.25.6367.
- [80] P. Vogl, H. P. Hjalmarson, and J. D. Dow. “A semi-empirical tight-binding theory of the electronic structure of semiconductors.” *J. Phys. Chem. Solids* **44**, 365 (1983). doi: 10.1016/0022-3697(83)90064-1.

- [81] G. W. Bryant. “Surface-enhanced alloy effects on $\text{Hg}_{1-x}\text{Cd}_x\text{Te}(100)$ surface states and the site-dependent coherent-potential approximation.” *Phys. Rev. B* **35**, 5547 (1987). doi: 10.1103/PhysRevB.35.5547.
- [82] J. C. Slater and G. F. Koster. “Simplified LCAO Method for the Periodic Potential Problem.” *Phys. Rev.* **94**, 1498 (1954). doi: 10.1103/PhysRev.94.1498.
- [83] L. C. Lew Yan Voon and M. Willatzen. *The $k \cdot p$ Method* (Springer-Verlag, 2009). ISBN 978-3-540-92871-3.
- [84] C. R. Becker, *et al.* “Band structure and its temperature dependence for type-III $\text{HgTe}/\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ superlattices and their semimetal constituent.” *Phys. Rev. B* **62**, 10353 (2000). doi: 10.1103/PhysRevB.62.10353.
- [85] K. Ortner, *et al.* “Valence band structure of $\text{HgTe}/\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ single quantum wells.” *Phys. Rev. B* **66**, 075322 (2002). doi: 10.1103/PhysRevB.66.075322.
- [86] C. R. Becker, V. Latussek, G. Landwehr, and L. W. Molenkamp. “Inverted band structure of type-III $\text{HgTe}/\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ superlattices and its temperature dependence.” *Phys. Rev. B* **68**, 035202 (2003). doi: 10.1103/PhysRevB.68.035202.
- [87] E. G. Novik, *et al.* “Band structure of semimagnetic $\text{Hg}_{1-y}\text{Mn}_y$ quantum wells.” *Phys. Rev. B* **72**, 035321 (2005). doi: 10.1103/PhysRevB.72.035321.
- [88] D.-W. Kim, *et al.* “Optoelectronic characteristics of $\text{CdTe}/\text{HgTe}/\text{CdTe}$ quantum-dot quantum-well nanoparticles.” *Sol. St. Comm.* **140**, 215 (2006). doi: 10.1016/j.ssc.2006.08.023.
- [89] C. R. Becker, *et al.* “Photoluminescence of $\text{HgTe}/\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ superlattices and a study of minibands.” *Phys. Rev. B* **75**, 115115 (2007). doi: 10.1103/PhysRevB.75.115115.
- [90] H. Ibach. *Physics of Surfaces and Interfaces* (Springer-Verlag, 2006). ISBN 978-3-540-34709-5.
- [91] R. Winkler, L. Y. Wang, Y. H. Lin, and C. S. Chu. “Robust Level Coincidences in the Subband Structure of Quasi 2D Systems.” arXiv:1011.4504 (2011).
- [92] M. von Truchseß, *et al.* “Magneto-optics and valence-band discontinuity in a $\text{HgTe}/\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ superlattice.” *Phys. Rev. B* **51**, 17618 (1995). doi: 10.1103/PhysRevB.51.17618.
- [93] M. von Truchseß, V. Latussek, C. Becker, and E. Batke. “Temperature dependent investigation of the HgTe/CdTe valence band offset.” *J. Cryst. Growth* **159**, 1128 (1996). doi: 10.1016/0022-0248(95)00757-1.
- [94] H. Mayer and U. Rössler. “Spin splitting and anisotropy of cyclotron resonance in the conduction band of GaAs.” *Phys. Rev. B* **44**, 9048 (1991). doi: 10.1103/PhysRevB.44.9048.
- [95] A. N. Redlich. “Parity violation and gauge noninvariance of the effective gauge field action in three dimensions.” *Phys. Rev. D* **29**, 2366 (1984). doi: 10.1103/PhysRevD.29.2366.
- [96] R. Jackiw. “Fractional charge and zero modes for planar systems in a magnetic field.” *Phys. Rev. D* **29**, 2375 (1984). doi: 10.1103/PhysRevD.29.2375.

- [97] A. A. Taskin and Y. Ando. “Quantum oscillations in a topological insulator $\text{Bi}_{1-x}\text{Sb}_x$.” *Phys. Rev. B* **80**, 085303 (2009). doi: 10.1103/PhysRevB.80.085303.
- [98] J. G. Checkelsky, *et al.* “Quantum Interference in Macroscopic Crystals of Nonmetallic Bi_2Se_3 .” *Phys. Rev. Lett.* **103**, 246601 (2009). doi: 10.1103/PhysRevLett.103.246601.
- [99] J. G. Analytis, *et al.* “Two-dimensional surface state in the quantum limit of a topological insulator.” *Nat. Phys.* **6**, 960 (2010). doi: 10.1038/nphys1861.
- [100] J. Xiong, *et al.* “Quantum oscillations in a topological insulator $\text{Bi}_2\text{Te}_2\text{Se}$ with large bulk resistivity.” *Phys. E* **44**, 917 (2012). doi: 10.1016/j.physe.2011.09.011.
- [101] T. B. Boykin. “Tight-binding-like expressions for the continuous-space electromagnetic coupling Hamiltonian.” *Am. J. Phys.* **69**, 793 (2001). doi: 10.1119/1.1344169.
- [102] J. Baars and F. Sorger. “Reststrahlen spectra of HgTe and $\text{Cd}_x\text{Hg}_{1-x}\text{Te}$.” *Sol. St. Comm.* **10**, 875 (1972). doi: 10.1016/0038-1098(72)90211-6.
- [103] M. C. Huang and P. Sikivie. “Structure of axionic domain walls.” *Phys. Rev. D* **32**, 1560 (1985). doi: 10.1103/PhysRevD.32.1560.
- [104] K. Lee. “Topological mass terms on axion domain walls.” *Phys. Rev. D* **35**, 3286 (1987). doi: 10.1103/PhysRevD.35.3286.
- [105] F. Wilczek. “Two applications of axion electrodynamics.” *Phys. Rev. Lett.* **58**, 1799 (1987). doi: 10.1103/PhysRevLett.58.1799.
- [106] R. D. Peccei and H. R. Quinn. “CP Conservation in the Presence of Pseudoparticles.” *Phys. Rev. Lett.* **38**, 1440 (1977). doi: 10.1103/PhysRevLett.38.1440.
- [107] P. Sikivie. “On the interaction of magnetic monopoles with axionic domain walls.” *Phys. Lett. B* **137**, 353 (1984). doi: 10.1016/0370-2693(84)91731-3.
- [108] C. G. Callan Jr., R. F. Dashen, and D. J. Gross. “The structure of the gauge theory vacuum.” *Phys. Lett. B* **63**, 334 (1976). doi: 10.1016/0370-2693(76)90277-X.
- [109] R. Jackiw and C. Rebbi. “Vacuum Periodicity in a Yang-Mills Quantum Theory.” *Phys. Rev. Lett.* **37**, 172 (1976). doi: 10.1103/PhysRevLett.37.172.
- [110] M. M. Vazifeh and M. Franz. “Quantization and 2π periodicity of the axion action in topological insulators.” *Phys. Rev. B* **82**, 233103 (2010). doi: 10.1103/PhysRevB.82.233103.
- [111] W.-K. Tse and A. H. MacDonald. “Giant Magneto-Optical Kerr Effect and Universal Faraday Effect in Thin-Film Topological Insulators.” *Phys. Rev. Lett.* **105**, 057401 (2010). doi: 10.1103/PhysRevLett.105.057401.
- [112] W.-K. Tse and A. H. MacDonald. “Magneto-optical and magnetoelectric effects of topological insulators in quantizing magnetic fields.” *Phys. Rev. B* **82**, 161104 (2010). doi: 10.1103/PhysRevB.82.161104.
- [113] L. Fu and C. L. Kane. “Superconducting Proximity Effect and Majorana Fermions at the Surface of a Topological Insulator.” *Phys. Rev. Lett.* **100**, 096407 (2008). doi: 10.1103/PhysRevLett.100.096407.

- [114] L. Fidkowski and A. Kitaev. “Effects of interactions on the topological classification of free fermion systems.” *Phys. Rev. B* **81**, 134509 (2010). doi: 10.1103/PhysRevB.81.134509.
- [115] D. A. Pesin and L. Balents. “Mott physics and band topology in materials with strong spin-orbit interaction.” *Nat. Phys.* **6**, 376 (2009). doi: 10.1038/nphys1606.
- [116] S. Rachel and K. Le Hur. “Topological insulators and Mott physics from the Hubbard interaction.” *Phys. Rev. B* **82**, 075106 (2010). doi: 10.1103/PhysRevB.82.075106.
- [117] M. Hohenadler, T. C. Lang, and F. F. Assaad. “Correlation Effects in Quantum Spin-Hall Insulators: A Quantum Monte Carlo Study.” *Phys. Rev. Lett.* **106**, 100403 (2011). doi: 10.1103/PhysRevLett.106.100403.
- [118] V. Gurarie. “Single-particle Green’s functions and interacting topological insulators.” *Phys. Rev. B* **83**, 085426 (2011). doi: 10.1103/PhysRevB.83.085426.
- [119] Y. Yamaji and M. Imada. “Mott physics on helical edges of two-dimensional topological insulators.” *Phys. Rev. B* **83**, 205122 (2011). doi: 10.1103/PhysRevB.83.205122.
- [120] S.-L. Yu, X. C. Xie, and J.-X. Li. “Mott Physics and Topological Phase Transition in Correlated Dirac Fermions.” *Phys. Rev. Lett.* **107**, 010401 (2011). doi: 10.1103/PhysRevLett.107.010401.
- [121] M. Hohenadler, *et al.* “Quantum phase transitions in the Kane-Mele-Hubbard model.” *Phys. Rev. B* **85**, 115132 (2012). doi: 10.1103/PhysRevB.85.115132.
- [122] V. N. Kotov, *et al.* “Electron-Electron Interactions in Graphene: Current Status and Perspectives.” arXiv:1012.3484 (2010).
- [123] S. Murakami. “Phase transition between the quantum spin Hall and insulator phases in 3D: emergence of a topological gapless phase.” *New J. Phys.* **9**, 356 (2007). doi: 10.1088/1367-2630/9/9/356.
- [124] S. Murakami, *et al.* “Tuning phase transition between quantum spin Hall and ordinary insulating phases.” *Phys. Rev. B* **76**, 205304 (2007). doi: 10.1103/PhysRevB.76.205304.
- [125] J. Lindhard. *On the properties of a gas of charged particles*, volume 28 of *Matematisk-fysiske meddelelser* (E. Munksgaard, 1954).
- [126] D. Pines and P. Nozières. *The Theory of Quantum Liquids*, volume 1 of *Advanced Book Classics* (Westview Press, 1999). ISBN 978-0-738-20229-7.
- [127] J. R. Schrieffer. *Theory of Superconductivity*. Advanced Books Classics, revised edition (Perseus Books, 1999). ISBN 978-0-738-20120-7.
- [128] M. Tinkham. *Introduction to Superconductivity*. Dover Books on Physics, second edition (Dover Books, 2004). ISBN 978-0-486-43503-9.
- [129] E. A. Andryushin and V. L. Ginzburg. *Superconductivity*. Revised edition (World Scientific Publishing Company, 2004). ISBN 978-9-812-38913-8.
- [130] A. Auerbach. *Interacting Electrons and Quantum Magnetism*. Graduate Texts in Contemporary Physics, second edition (Springer-Verlag, 1998). ISBN 978-0-387-94286-5.

- [131] P. M. Chaikin and T. C. Lubensky. *Principles of Condensed Matter Physics* (Cambridge University Press, 2000). ISBN 978-0-521-79450-3.
- [132] G. Grüner. *Density Waves in Solids*. Frontiers in Physics (Westview Press, 2000). ISBN 978-0-738-20304-1.
- [133] E. Wigner. “On the Interaction of Electrons in Metals.” *Phys. Rev.* **46**, 1002 (1934). doi: 10.1103/PhysRev.46.1002.
- [134] D. M. Ceperley and B. J. Alder. “Ground State of the Electron Gas by a Stochastic Method.” *Phys. Rev. Lett.* **45**, 566 (1980). doi: 10.1103/PhysRevLett.45.566.
- [135] B. Tanatar and D. M. Ceperley. “Ground state of the two-dimensional electron gas.” *Phys. Rev. B* **39**, 5005 (1989). doi: 10.1103/PhysRevB.39.5005.
- [136] K. S. Novoselov, *et al.* “Electric Field Effect in Atomically Thin Carbon Films.” *Science* **306**, 666 (2004). doi: 10.1126/science.1102896.
- [137] J. González, F. Guinea, and M. A. H. Vozmediano. “Non-Fermi liquid behavior of electrons in the half-filled honeycomb lattice (A renormalization group approach).” *Nucl. Phys. B* **424**, 595 (1994). doi: 10.1016/0550-3213(94)90410-3.
- [138] D. V. Khveshchenko. “Massive Dirac fermions in single-layer graphene.” *J. Phys.: Condens. Matt.* **21**, 075303 (2009). doi: 10.1088/0953-8984/21/7/075303.
- [139] G.-Z. Liu, W. Li, and G. Cheng. “Interaction and excitonic insulating transition in graphene.” *Phys. Rev. B* **79**, 205429 (2009). doi: 10.1103/PhysRevB.79.205429.
- [140] O. V. Gamayun, E. V. Gorbar, and V. P. Gusynin. “Gap generation and semimetal-insulator phase transition in graphene.” *Phys. Rev. B* **81**, 075429 (2010). doi: 10.1103/PhysRevB.81.075429.
- [141] J.-R. Wang and G.-Z. Liu. “Absence of dynamical gap generation in suspended graphene.” *New J. Phys.* **14**, 043036 (2012).
- [142] S. P. Klevansky and R. H. Lemmer. “Chiral-symmetry breaking at finite temperatures.” *Phys. Rev. D* **38**, 3559 (1988). doi: 10.1103/PhysRevD.38.3559.
- [143] C. D. Roberts and B. H. J. McKellar. “Critical coupling for dynamical chiral-symmetry breaking.” *Phys. Rev. D* **41**, 672 (1990). doi: 10.1103/PhysRevD.41.672.
- [144] R. M. Woloshyn and F. X. Lee. “Chiral symmetry breaking and cooling in lattice QCD.” *Phys. Rev. D* **53**, 1709 (1996). doi: 10.1103/PhysRevD.53.1709.
- [145] A. P. Szczepaniak and P. Krupinski. “Spontaneous chiral symmetry breaking in the linked cluster expansion.” *Phys. Rev. D* **66**, 096006 (2002). doi: 10.1103/PhysRevD.66.096006.
- [146] B. Jain, I. Mitra, and H. S. Sharatchandra. “Criterion for dynamical chiral symmetry breaking.” *Phys. Rev. D* **76**, 025020 (2007). doi: 10.1103/PhysRevD.76.025020.
- [147] A. C. Aguilar and J. Papavassiliou. “Chiral symmetry breaking with lattice propagators.” *Phys. Rev. D* **83**, 014013 (2011). doi: 10.1103/PhysRevD.83.014013.
- [148] R. D. Pisarski. “Chiral-symmetry breaking in three-dimensional electrodynamics.” *Phys. Rev. D* **29**, 2423 (1984). doi: 10.1103/PhysRevD.29.2423.

- [149] K. Higashijima. “Dynamical chiral-symmetry breaking.” *Phys. Rev. D* **29**, 1228 (1984). doi: 10.1103/PhysRevD.29.1228.
- [150] T. W. Appelquist, M. Bowick, D. Karabali, and L. C. R. Wijewardhana. “Spontaneous chiral-symmetry breaking in three-dimensional QED.” *Phys. Rev. D* **33**, 3704 (1986). doi: 10.1103/PhysRevD.33.3704.
- [151] D. C. Elias, *et al.* “Dirac cones reshaped by interaction effects in suspended graphene.” *Nat. Phys.* **7**, 701 (2011). doi: 10.1038/nphys2049.
- [152] M. S. Dresselhaus, G. Dresselhaus, and J. Ado. *Many-Particle Physics*. First edition (Springer-Verlag, 2008). ISBN 978-3-540-32897-1.
- [153] H. Arnold, *et al.* *International Tables for Crystallography*, volume A. Fifth edition (Kluwer Academic Publishers, 2005). ISBN 0-7923-6590-9.
- [154] P. R. Wallace. “The Band Theory of Graphite.” *Phys. Rev.* **71**, 622 (1947). doi: 10.1103/PhysRev.71.622.
- [155] A. H. Castro Neto, *et al.* “The electronic properties of graphene.” *Rev. Mod. Phys.* **81**, 109 (2009). doi: 10.1103/RevModPhys.81.109.
- [156] N. M. R. Peres. “*Colloquium*: The transport properties of graphene: An introduction.” *Rev. Mod. Phys.* **82**, 2673 (2010). doi: 10.1103/RevModPhys.82.2673.
- [157] S. Das Sarma, S. Adam, E. H. Hwang, and E. Rossi. “Electronic transport in two-dimensional graphene.” *Rev. Mod. Phys.* **83**, 407 (2011). doi: 10.1103/RevModPhys.83.407.
- [158] M. O. Goerbig. “Electronic properties of graphene in a strong magnetic field.” *Rev. Mod. Phys.* **83**, 1193 (2011). doi: 10.1103/RevModPhys.83.1193.
- [159] D. Gosálbez-Martínez, D. Soriano, J. J. Palacios, and J. Fernández-Rossier. “Spin-filtered edge states in graphene.” arXiv:1203.2479 (2012).
- [160] M. Fujita, K. Wakabayashi, K. Nakada, and K. Kusakabe. “Peculiar Localized State at Zigzag Graphite Edge.” *J. Phys. Soc. Jpn.* **65**, 1920 (1996). doi: 10.1143/JPSJ.65.1920.
- [161] K. Nakada, M. Fujita, G. Dresselhaus, and M. S. Dresselhaus. “Edge state in graphene ribbons: Nanometer size effect and edge shape dependence.” *Phys. Rev. B* **54**, 17954 (1996). doi: 10.1103/PhysRevB.54.17954.
- [162] A. H. Castro Neto, F. Guinea, and N. M. R. Peres. “Edge and surface states in the quantum Hall effect in graphene.” *Phys. Rev. B* **73**, 205408 (2006). doi: 10.1103/PhysRevB.73.205408.
- [163] Y.-W. Son, M. L. Cohen, and S. G. Louie. “Half-metallic graphene nanoribbons.” *Nature* **444**, 347 (2006). doi: 10.1038/nature05180.
- [164] Y.-W. Son, M. L. Cohen, and S. G. Louie. “Energy Gaps in Graphene Nanoribbons.” *Phys. Rev. Lett.* **97**, 216803 (2006). doi: 10.1103/PhysRevLett.97.216803.
- [165] J. Fernández-Rossier and J. J. Palacios. “Magnetism in Graphene Nanoislands.” *Phys. Rev. Lett.* **99**, 177204 (2007). doi: 10.1103/PhysRevLett.99.177204.

- [166] K. Sasaki and R. Saito. “Magnetism as a Mass Term of the Edge States in Graphene.” *J. Phys. Soc. Jpn.* **77**, 054703 (2008). doi: 10.1143/JPSJ.77.054703.
- [167] O. Hod, V. Barone, and G. E. Scuseria. “Half-metallic graphene nanodots: A comprehensive first-principles theoretical study.” *Phys. Rev. B* **77**, 035411 (2008). doi: 10.1103/PhysRevB.77.035411.
- [168] O. V. Yazyev and M. I. Katsnelson. “Magnetic Correlations at Graphene Edges: Basis for Novel Spintronics Devices.” *Phys. Rev. Lett.* **100**, 047209 (2008). doi: 10.1103/PhysRevLett.100.047209.
- [169] A. Mañanes, *et al.* “Half-metallic finite zigzag single-walled carbon nanotubes from first principles.” *Phys. Rev. B* **78**, 035432 (2008). doi: 10.1103/PhysRevB.78.035432.
- [170] K. Harigaya. “The mechanism of magnetism in stacked nanographite: theoretical study.” *J. Phys.: Condens. Matt.* **13**, 1295 (2001).
- [171] K. Harigaya and T. Enoki. “Mechanism of magnetism in stacked nanographite with open shell electrons.” *Chem. Phys. Lett.* **351**, 128 (2002). doi: 10.1016/S0009-2614(01)01338-0.
- [172] B. Wunsch, T. Stauber, F. Sols, and F. Guinea. “Dynamical polarization of graphene at finite doping.” *N. J. Phys.* **8**, 318 (2006).
- [173] B. Wunsch, T. Stauber, and F. Guinea. “Electron-electron interactions and charging effects in graphene quantum dots.” *Phys. Rev. B* **77**, 035316 (2008). doi: 10.1103/PhysRevB.77.035316.
- [174] B. Wunsch, T. Stauber, F. Sols, and F. Guinea. “Interactions and Magnetism in Graphene Boundary States.” *Phys. Rev. Lett.* **101**, 036803 (2008). doi: 10.1103/PhysRevLett.101.036803.
- [175] T. Giamarchi. *Quantum Physics in One Dimension* (Oxford University Press, 2003). ISBN 978-0-198-52500-4.
- [176] G. E. Volovik. *The Universe in a Helium Droplet*, volume 117 of *International Series of Monographs on Physics* (Oxford University Press, 2009). ISBN 978-0-199-56484-2.
- [177] G. E. Volovik. “Flat band in topological matter: possible route to room-temperature superconductivity.” arXiv:1110.4469 (2011).
- [178] T. Heikkilä and G. Volovik. “Dimensional crossover in topological matter: Evolution of the multiple Dirac point in the layered system to the flat band on the surface.” *JETP Letters* **93**, 59 (2011). doi: 10.1134/S002136401102007X.
- [179] T. Heikkilä, N. Kopnin, and G. Volovik. “Flat bands in topological media.” *JETP Letters* **94**, 233 (2011). doi: 10.1134/S0021364011150045.
- [180] C. Fang, M. J. Gilbert, X. Dai, and B. A. Bernevig. “Double-Weyl Topological Metals Stabilized by Point Group Symmetry.” arXiv:1111.7309 (2011).
- [181] L. Balents. “Viewpoint: Weyl electrons kiss.” *Physics* **4** (2011). doi: 10.1103/Physics.4.36.
- [182] A. A. Burkov and L. Balents. “Weyl Semimetal in a Topological Insulator Multilayer.” *Phys. Rev. Lett.* **107**, 127205 (2011). doi: 10.1103/PhysRevLett.107.127205.

- [183] W. P. Su, J. R. Schrieffer, and A. J. Heeger. "Solitons in Polyacetylene." *Phys. Rev. Lett.* **42**, 1698 (1979). doi: 10.1103/PhysRevLett.42.1698.
- [184] W. P. Su, J. R. Schrieffer, and A. J. Heeger. "Soliton excitations in polyacetylene." *Phys. Rev. B* **22**, 2099 (1980). doi: 10.1103/PhysRevB.22.2099.
- [185] A. J. Heeger, S. Kivelson, J. R. Schrieffer, and W. P. Su. "Solitons in conducting polymers." *Rev. Mod. Phys.* **60**, 781 (1988). doi: 10.1103/RevModPhys.60.781.
- [186] R. Jackiw and C. Rebbi. "Solitons with fermion number $1/2$." *Phys. Rev. D* **13**, 3398 (1976). doi: 10.1103/PhysRevD.13.3398.

Eigenständigkeitserklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Hr. Prof. Dr. Achim Rosch betreut worden.

Köln, den 25.11.2012

Teilpublikation:

Matthias Sitte, Achim Rosch, Ehud Altman, and Lars Fritz: “Topological Insulators in Magnetic Fields: Quantum Hall Effect and Edge Channels with Nonquantized θ Term.” *Phys. Rev. Lett.* **108**, 126807 (2012). doi: 10.1103/PhysRevLett.108.126807.

Lebenslauf

Persönliche Daten

Name	Matthias Sitte
Geburtsdatum	03.12.1982
Geburtsort	Olpe
Familienstand	verheiratet
Staatsangehörigkeit	deutsch

Schulbildung und Wehrdienst

07/1989 – 06/1993	Gemeinschaftsgrundschule Geilenkircher Str., Köln
07/1993 – 06/2002	Städtisches Gymnasium Kreuzgasse, Köln
06/2002	Abitur
07/2002 – 03/2003	Grundwehrdienst im Fernmeldeaufklärungsregiment 940, Daun

Hochschulstudium

04/2003 – 07/2008	Studium der Physik an der Universität zu Köln
11/2003 – 07/2008	Stipendium der Studienstiftung des Deutschen Volkes
04/2004 – 07/2008	Studium der Mathematik an der Universität zu Köln
04/2007 – 10/2012	Studentisches Mitglied der Bonn-Cologne Graduate School
07/2008	Diplom in Physik
07/2008 – 10/2012	Promotion in Theoretischer Physik

Arbeitsverhältnisse

10/2004 – 07/2005	Studentische Hilfskraft am Mathematischen Institut der Universität zu Köln
10/2006 – 07/2008	Studentische Hilfskraft am Institut für Theoretische Physik der Universität zu Köln
07/2008 – heute	Wissenschaftlicher Mitarbeiter (ibidem)

