Summer 8-12-2011

Topological Insulators and Superconductors

Chi Yan Jeffrey Teo
h0105889@gmail.com

Topological Insulators and Superconductors

Abstract

We study theoretical properties of robust low energy electronic excitations associated with topological insulators and superconductors. The bulk materials are described by non-interacting single particle band Hamiltonians with a finite excitation gap. Their topological phases are classifed according to symmetries and dimensions, characterized by discrete bulk invariants, and correspond to topologically protected gapless excitations bounded to boundaries, interfaces or other kinds of defects. In particular, we study the metallic surface states of the three dimensional topological insulator $Bi_{1-x}Sb_x$, critical edge transport behavior of quantum spin Hall insulators (QSHI) using point contact geometry, Majorana bound states in three dimensions and their resemblance to Ising statistics, and various gapless modes accompanying topological defects in insulators and superconductors. We illustrate the topological phase of $Bi_{1-x}Sb_x$ by calculating its surface energy spectrum numerically from a previously proposed tight binding model. An odd number of surface Dirac cones occupy the surface Brillouin zone and exhibit the strong topological nature of the material. We investigate the critical conductance behavior of a point contact in QSHI using a spinful Luttinger liquid description along the edges. For weak interactions, a novel intermediate fixed point controls the pinch-off transition, and the universal crossover scaling function of conductance is extracted from the solvable limits for the Luttinger parameter $g=1-\varepsilon$, $g=1/2+\varepsilon$, and $g=1/3^{1/2}$. Majorana fermions are studied as zero energy quasiparticle excitations associated with pointlike topological defects in 3D superconductors. The low energy modes are described phenomenologically in a Dirac-type Bogoliubov de Gennes (BdG) framework, and the Majorana bound states are shown to exhibit Ising non-Abelian statistics despite living in (3+1) dimensions. In particular, novel braidless operations are shown to be responsible for fermion parity pumping processes, and are unique features in 3D. A unified framework to classify topological defects in insulators and superconductors is developed. A 2+8-fold periodic classification is discovered. A generalized bulk-boundary correspondence equates the topology to robust gapless defect modes. Physical proposals are made especially using heterostructures to achieve desirable low energy electronic excitations in line and point defects as well as adiabatic cycles.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Physics & Astronomy

First Advisor
Charles L. Kane

Keywords
electronic band theory, topological defect

Subject Categories
Condensed Matter Physics

TOPOLOGICAL INSULATORS AND SUPERCONDUCTORS

Jeffrey C.Y. Teo

A DISSERTATION

in

Physics and Astronomy

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2011

_____
Charles L. Kane, Professor of Physics and Astronomy
Supervisor of Dissertation

_____
A.T. Charlie Johnson, Professor of Physics and Astronomy
Graduate Group Chairperson

Dissertation Committee
Randall D. Kamien, Professor of Physics and Astronomy
Justin Khoury, Assistant Professor of Physics and Astronomy
Jay Kikkawa, Associate Professor of Physics and Astronomy
Eugene J. Mele, Professor of Physics and Astronomy

Topological Insulators and Superconductors

COPYRIGHT

Chi Yan Jeffrey Teo

2011

Acknowledgements

This dissertation is based on the research I have done in the previous five years under the supervision of Charlie Kane. I would like to express my gratitude to him for his guidance and education. I learned a tremendous amount of physical knowledge in the process of working with him as well as during attending his lectures. Perhaps more importantly, he taught me the ways to tackle a research problem by simple but sophisticated understanding of the physical issues, rather than dwelling on complicated mathematics that could turn out unnecessary or even irrelevant. I greatly appreciate his patience in explaining or discussing physical topics. He is always willing to spend time in sharing valuable knowledge and insights. I am also very grateful for being able to travel a lot and participate in many conferences and workshops with him.

Among my professors, I would especially like to thank Randy Kamien and Gene Mele for helping me in my postdoctoral application. I always enjoyed Randy's lecture. They were very interactive and stimulating. I benefited a lot from the Friday lunch meetings with Charlie, Gene and Andrew Rappe. They provided the opportunity for me to regularly communicate with a group of physicists and

ABSTRACT

TOPOLOGICAL INSULATORS AND SUPERCONDUCTORS

Jeffrey C.Y. Teo

Charles L. Kane

We study theoretical properties of robust low energy electronic excitations associated with topological insulators and superconductors. The bulk materials are described by non-interacting single particle band Hamiltonians with a finite excitation gap. Their topological phases are classifed according to symmetries and dimensions, characterized by discrete bulk invariants, and correspond to topologically protected gapless excitations bounded to boundaries, interfaces or other kinds of defects. In particular, we study the metallic surface states of the three dimensional topological insulator $Bi_{1-x}Sb_x$, critical edge transport behavior of quantum spin Hall insulators (QSHI) using point contact geometry, Majorana bound states in three dimensions and their resemblance to Ising statistics, and various gapless modes accompanying topological defects in insulators and superconductors.

We illustrate the topological phase of $Bi_{1-x}Sb_x$ by calculating its surface energy spectrum numerically from a previously proposed tight binding model. An odd number of surface Dirac cones occupy the surface Brillouin zone and exhibit the strong topological nature of the material. We investigate the critical conductance behavior of a point contact in QSHI using a spinful Luttinger liquid description along the edges. For weak interactions, a novel intermediate fixed point controls the

pinch-off transition, and the universal crossover scaling function of conductance is extracted from the solvable limits for the Luttinger parameter $g = 1-\epsilon$, $g = 1/2+\epsilon$, and $g = 1/\sqrt{3}$.

Majorana fermions are studied as zero energy quasiparticle excitations associated with pointlike topological defects in 3D superconductors. The low energy modes are described phenomenologically in a Dirac-type Bogoliubov de Gennes (BdG) framework, and the Majorana bound states are shown to exhibit Ising non-Abelian statistics despite living in $(3 + 1)$ dimensions. In particular, novel braidless operations are shown to be responsible for fermion parity pumping processes, and are unique features in 3D.

A unified framework to classify topological defects in insulators and superconductors is developed. A $2+8$-fold periodic classification is discovered. A generalized bulk-boundary correspondence equates the topology to robust gapless defect modes. Physical proposals are made especially using heterostructures to achieve desirable low energy electronic excitations in line and point defects as well as adiabatic cycles.

Contents

List of Tables

List of Figures

Preface

This dissertation is based heavily on the four projects I contributed to during the period of my Ph.D. program.

1. *Surface States of the Topological Insulator $Bi_{1-x}Sb_x$* [Teo, Fu and Kane, 2008],

2. *Critical Behavior of a Point Contact in a Quantum Spin Hall Insulator* [Teo and Kane, 2009],

3. *Majorana Fermions and Non-Abelian Statistics in Three Dimensions* [Teo and Kane, 2010a], and

4. *Topological Defects and Gapless Modes in Insulators and Superconductors* [Teo and Kane, 2010b].

The published articles mentioned above are attached as appendices. To avoid repetition, this dissertation will focus more on the results and conclusions instead of providing detailed steps of calculation. The methodology will be described but not carried out explicitly. Some omitted topics in the published articles will also be discussed.

Section 1.1 is an introduction on single particle topological phases of matter, including quantum Hall effect and topological insulators. Section 4.1 provides the theoretical background for Majorana fermions in various condensed matter models. Section 5.1 contains previous results on the classifcation of topological insulators and superconductors. These reviews from the literature that lead to the four main subjects of the dissertation. Contents in the other sections are mostly originated from the four articles unless specifically referred otherwise.

In the study of bismuth antimony (chapter 2), we contributed to the surface spectrum calculation (section 2.1), the analysis on inversion symmetry and surface fermion parity (section 2.2) and the mirror eigenvalues verification that leads to the anomalous mirror chirality (section 2.3). For QSHI point contact (chapter 3), we provided perturbative RG analysis on the critical conductance at the solvable regions $g = 1 - \epsilon$, $g = 1/2 + \epsilon$ and the self-dual fixed point at $g = 1/\sqrt{3}$. For 3D Majorana fermions (chapter 4) and topological defects (chapter 5), we developed topological characterizations, such as the various topological invariants in sections 4.2, 4.3.1 and 5.2 and came up with the braidless-exchange operations in sections 4.3.2-4.3.3, and the defect models in section 5.3. All of my contributions are results of consultations with my advisor and fellow colleagues and none of them would be possible without their help and support.

Chapter 1

Introduction

One of the greatest triumphs of physics in modern history is the discovery of quantum Hall effect [Klitzing, Dorda and Pepper, 1980]. The quantized Hall conductivity is measured with remarkable accuracy ($\sim$1 part in 10^9) and sets the standard for the universal constant e^2/h of conductivity. Surprisingly, the integer Hall plateaus are measured by relatively dirty samples of semiconductors. This rules out the possibility of broken symmetry phases as there are no order parameters in the system. Quantum Hall effect thus gives rise to a new generation of phases of matter, known as *topological phases* [Thouless *et.al.*, 1982; Wen, 1995]. They are characterized by the topology of their ground state, in the sense that certain quantities (conductivity, charge and statistics of excitations, ground state degeneracies, etc) are invariant under small perturbations (such as disorder and interaction) and cannot change without closing the energy gap of excitations in the bulk of the material.

All quantum Hall phases are two dimensional time reversal breaking phases.

Electrons are effectively trapped on a 2D plane such as a stratified GaAs heterostructure or a layer of graphene, and experience a magnetic field perpendicular to the surface. For a semiconductor with $\sim 10^{12}$ number of charge carriers per cm^2, one typically needs a field of $\sim 10T$ to observe the first few Hall plateaus. The energy scale is set by the cyclotron frequency, which then falls in the temperature scale of a few Kelvin. All these are probably not the most appealing aspects of quantum Hall physics to a semiconductor engineer who works in room environment.

Roughly a quarter of a century after the discovery of quantum Hall effect, a new category of topological phases was proposed theoretically [Kane and Mele, 2005a,b; Fu, Kane and Mele, 2007; Moore and Balents, 2007; Roy, 2006b] then followed swiftly by experimental realization [König *et.al.*, 2007; Hsieh *et.al.*, 2008; Xia *et al.*, 2009a]. These are called *topological insulators* (TI) [Qi and Zhang, 2010a; Moore, 2010; Hasan and Kane, 2010; Kane and Moore, 2011]. They are topological phases that exist in both two and three dimensional spin-orbit coupled materials, preserve time reversal symmetry and hence do not require an external field, and occur in room temperature bulk insulators (band gap up to ~ 300meV).

Similar to a quantum Hall insulator, a topological insulator necessarily carries metallic gapless boundary excitations. The boundary energy spectrum can be experimentally observed via angle resolved photoemission spectroscopy (ARPES) and forms the signature of this type of topological phases.[1] The equality between the topological characteristics of a bulk gapped material and its gapless boundary states is known as *bulk-boundary correspondence*. In particular, the topological nature of the bulk is measured by the violation of fermion doubling theorem [Nielssen and

---

[1]See definition 1.1.1 and remarks thereafter.

Ninomiya, 1983] on the boundary, i.e. the boundary states cannot be reproduced by any pure $(d-1)$-dimensional lattice fermion theory.

The bulk-boundary correspondence will be the main theme of this thesis. Although it could be mathematically reformulated as a generalization to the Atiyah-Singer index theorem [Atiyah and Singer, 1963, 1968; Atiyah and Segal, 1968], a rigorous proof of that inevitably requires the language of topological K-theory as well as some heavy machinery in differential operator theory. These would be beyond the scope of this dissertation. Physical argument will however be used to affirm the correspondence. In particular, bulk-boundary correspondence has been proven or verified in many special cases such as integer quantum Hall effect [Volovik, 2003], Su-Schrieffer-Heeger model [Su, Schrieffer and Heeger, 1980], Kitaev's Majorana chain [Kitaev, 2001], Jackiw-Rossi model [Jackiw and Rossi, 1981], and many more.

The main focus of this thesis is to explore the physical nature and implications of robust boundary gapless excitations [Teo, Fu and Kane, 2008; Teo and Kane, 2009, 2010a,b]. A review on topological insulators will first be given in section 1.1. It will cover the basic ideas of integer quantum Hall effects [Thouless *et.al.*, 1982; Halperin, 1982; Haldane, 1988] that leads to the study of topology of valence states. The theory of topological insulators in two and three dimensions will be reviewed [Kane and Mele, 2005a,b; Fu, Kane and Mele, 2007; Moore and Balents, 2007; Fu and Kane, 2007; Roy, 2006a,b; Qi, Hughes and Zhang, 2008]. $\mathbb{Z}_2$-topological invariants will be introduced to characterize a time reversal symmetric spin-orbit coupled insulator and related to robust gapless boundary excitations.

It will be followed by the study of surface states of a 3D topological insulator [Teo, Fu and Kane, 2008]. A tight binding model generated surface spectrum of the 3D material bismuth-antimony ($Bi_{1-x}Sb_x$) alloy will be shown in chapter 2. It carries the signature of a strong topological insulator by exhibiting an odd number of Dirac cones on the surface Brillouin zone. The surface spectrum structure of a general inversion symmetric insulator will be investigated using *surface Fermion parity*. A new invariant known as *Mirror chirality* η_M will be introduced to explain the band crossings and hole pockets in the surface excitations.

Chapter 3 will focus on electrical transport behavior of a quantum spin Hall insulator using a quantum point contact [Teo and Kane, 2009]. The effect of electron-electron interactions on the gapless edge modes of a 2D quantum spin Hall insulator can be qualitatively captured in a spinful Luttinger liquid (SLL) description [Kane and Fisher, 1992a; Hou *et.al.*, 2009]. In the weak interaction limit $1/2 < g < 2$, it will be shown that the stable phases of the point contact are either completely closed or open. The pinch-off transition behavior is determined by a novel intermediate fixed point. It controls the critical behavior of conductance, which collapses onto a single universal scaling function capturing its dependence on temperature and gate voltage. These can be analytically studied at the solvabale regimes $g = 1 - \epsilon$, $g = 1/2 + \epsilon$, and $g = 1/\sqrt{3}$.

The massless surface state of a topological insulator (TI) can be gapped by breaking either time reversal or charge conservation symmetry. Chapter 4 will focus on the consequences of the latter. A pairing gap can be introduced due to proximity with a superconductor (SC) [Fu and Kane, 2008]. This provides a new platform for

realizing Majorana fermions in condensed matter systems. Although the Majorana excitations at a TI-SC heterostructure can be understood by an effective surface theory, Majorana fermions actually live in 3D along a smooth TI-SC transition interface. Therefore, unlike a pure 2D topological phase such as the non-Abelian 5/2 fractional quantum Hall state [Moore and Read, 1991; Nayak *et.al.*, 2008], p-wave superconductor [Das Sarma, Nayak and Tewari, 2006] or the Kitaev honeycomb model [Kitaev, 2006], a 3D theory is essential for a satisfactory understanding of Majorana fermions in heterostructures [Teo and Kane, 2010a].

Section 4.2 will begin by explaining the topological origin of Majorana fermions as excitations at a point defect. A $\mathbb{Z}_2$ invariant will be introduced based on the modulation of a Bogoliubov de Gennes (BdG) Hamiltonian around the defect. The existence of Majorana zero modes will be illustrated by an exact solvable minimal Dirac model. The phase of the zero mode will be controlled by the mass term of the Dirac Hamiltonian. The configuration of the mass term is a non-local degree of freedom originating from the gapped BdG Hamiltonian throughout the entire 3D system, and will be discussed in section 4.3. This allows the possibility of realizing Ising anyonic statistics in three dimensions and new kinds of operations that do not rely on braiding of physical positions of Majorana fermions.

After the discussion on the boundary states of topological insulators and Majorana zero modes of superconductors, chapter 5 will focus on describing a general framework to classify topological defects in insulators and superconductors, and their associated robust gapless excitations [Teo and Kane, 2010b]. An introduction on the 2+8 fold periodic topological classification of band Hamiltonians with

full crystalline translational symmetries will be given [Schnyder *et.al.*, 2008, 2009; Kitaev, 2009], resembling the Bott periodicity in stable homotopy groups of symmetric spaces [Cartan, 1926; Bott, 1956]. It will be followed by the generalization to topological defects in section 5.2. Topological invariants will be constructed in the form of explicit integral formulae that could be used to determine the topological phases of any given band theory. A generalization of bulk-boundary correspondence will be introduced in section 5.3 and this will relate the topological classes of defect Hamiltonians to the presence of protected gapless modes at defects. An abundant amount of old and new examples of point defects, line defects and temporal cycles will be studied. Many of them exhibit desirable electronic transport properties including chiral/helical Dirac/Majorana modes in three dimensions.

1.1 Single particle topological insulating phases

Many electronic properties of a semiconductor can be captured by an effective single particle non-interacting Hamiltonian [Bloch, 1929]. In principle, this can be constructed via a pseudo-potential approximation in the free electron limit [Phillips and Kleinman, 1959], an LCAO approximation in the tight binding limit [Slater and Koster, 1954], or a density functional calculation [Kohn and Sham, 1965]. In a crystalline material, the Hamiltonian respects discrete translational symmetry and energy states are labeled by crystal momentum $\mathbf{k}$ which lives in a toric Brilloun zone $T^d = U(1)^d$. The Bloch Hamiltonian $\mathcal{H}(\mathbf{k})$ is a matrix, whose eigenstates $|u_m(\mathbf{k})\rangle$ are called Bloch states and momentum labeled eigenvalues $E_m(\mathbf{k})$ are called energy bands.

An insulator is described by a gapped band Hamiltonian with a non-zero energy gap separating the occupied valence bands and unoccupied conduction bands. While the kinematics of quasi-particles, such as the effective mass tensor, are described by the dispersion of the band spectrum, the topology of an insulator is encoded in its Bloch states [Thouless *et.al.*, 1982]. It reflects certain robust properties of the material that cannot be changed without going through a metallic transition that closes the energy gap. A more precise notion of a topological phase can be formulated as the following.

Definition 1.1.1. (Topological equivalence) *Two gapped band Hamiltonians are topologically equivalent if they can be smoothly deformed into one another while maintaining the energy gap and respecting the prescribed symmetries (time reversal, particle-hole, sublattice, inversion, rotation, mirror, etc.).*

Moreover, two band models should be treated as topologically equivalent if they differ from each other only by the number of trivial low lying or high energy bands, such as core atomic orbitals or energetically irrelevant excitations. These models will be treated as *stably* topologically equivalent, and the collection of such an equivalent class of Hamiltonians will be referred in this thesis as a *topological phase*. The study of topological phases of gapped band Hamiltonians is called *topological band theory*, and applies not only to insulators but also to superconductors [Hasan and Kane, 2010; Qi and Zhang, 2010b].

Remarks on terminology

1. Not all topological phases carry robust gapless boundary excitations, especially

those associated with non-local point group symmetries that would be broken by spatially terminating the material. These phases will not be addressed in this dissertation.

2. Although formulated as a theory on a lattice, translation symmetry is not absolutely essential in a topological phase of matter. One could in fact extend definition 1.1.1 to incorporate weak disorder using the language of C^*-algebras [Bellissard, van Elst and Schulz-Baldes, 1994] or almost commuting matrices [Hasting and Loring, 2010]. Topological invariants could be written down but would involve cumbersome expressions of certain *trace-class* operators. These would not be covered in the thesis, however only topological phases that are stable under weak disorder will be considered.

3. The notion of topological phase in this dissertation concerns single particle non-interacting Hamiltonians only. It should not be confused with *topological order* in a many-body interacting system [Wen, 1995] that indicates other topological information, such as the ground state degeneracy [Wen and Niu, 1990] and fusion properties of excitations [Moore and Read, 1991; Read and Rezayi, 1999; Kitaev, 2006], which are only relevant in a many-body context.

Integer quantum Hall states are topologcal insulating phases that break time reversal symmetry, while topological insulators are time reversal symmetric topological phases. Their topologies rely on the non-trivialities of occupied states or valence bands in the bulk, and are revealed by gapless excitations that are robust against backscattering and localization. These properties are going to be reviewed below.

1.1.1 Integer quantum Hall effect

An analytic derivation of the quantized Hall conductivity in a periodic potential was given by Thouless, Kohmoto, Nightingale and den Nijs (TKNN) in 1982. They considered a two dimensional electron gas (2DEG) under a uniform perpendicular magnetic field and an arbitrary commensurating[2] lattice potential described by a single particle Hamiltonian. The conductivity was expressed using the Kubo formula of linear response as an integral involving valence band states.

$$\sigma_{xy} = \nu \frac{e^2}{h} \tag{1.1.1}$$

$$\nu = \frac{i}{2\pi} \int_{BZ} d^2k \sum_{E_m < E_f} 2i \mathrm{Im} \left\langle \frac{\partial}{\partial k_x} u_m(\mathbf{k}) \middle| \frac{\partial}{\partial k_y} u_m(\mathbf{k}) \right\rangle \tag{1.1.2}$$

$$= \frac{i}{2\pi} \int_{BZ} \mathrm{Tr}\left(\mathcal{F}\right) \in \mathbb{Z} \tag{1.1.3}$$

Here the summation is taken over the occupied electron subbands and the integral is over the magnetic Brillouin zone (BZ) [Zak, 1964]. And $\mathcal{F}$ is Berry curvature [Berry, 1984] defined by the Cartan's structure formula

$$\mathcal{F} = d\mathcal{A} + \mathcal{A} \wedge \mathcal{A} \tag{1.1.4}$$

associated to the Berry connection

$$\mathcal{A}_{mn}(\mathbf{k}) = \langle u_m(\mathbf{k}) | du_n(\mathbf{k}) \rangle \tag{1.1.5}$$

---

[2]The magnetic flux per unit cell is assumed to be an integer multiple of the magnetic flux quantum $\phi_0 = hc/e$. This is to ensure a non-zero excitation gap in the non-interactng theory, and to guarantee the commutativity of the discrete magnetic translation group [Zak, 1964] so that a magnetic Brillouin zone can be defined.

constructed from occupied subbands $u_m(\mathbf{k})$.

The topology of occupied bands is characterized by the TKNN integer ν, also known as the first Chern number[3] [Chern, 1946; Nakahara, 1990]. This integral invariant carries only information about the occupied states, and therefore is insensative to perturbations that alter the energy spectrum of the system. Being discrete, it is impossible for the number to change upon a continuous deformation of the system. Eq.(1.1.2) is invalid only when the system is undergoing a quantum phase transition where the excitation gap vanishes. Hence, the TKNN integer labels a topological phase of matter according to definition 1.1.1.

The integer quantum Hall effect can also be understood by looking at edge excitations [Halperin, 1982]. Under a perpendicular magnetic field, a 2DEG splits into Landau levels with spatially separated linear momentum modes. The TKNN integer is given by the filling fraction, which is the number of filled Landau levels.

$$\nu = \frac{N_e}{N_B} = \frac{\text{number of electrons}}{\text{number of magnetic flux quanta}} \qquad (1.1.6)$$

Mid-gap excitations arise due to band bending at the boundary (figure 1.1), and these are unidirectional modes exponentially localized along the edges known as chiral modes. The integer quantum Hall effect can also be mimicked by a time reversal breaking tight binding model on the graphene lattice [Haldane, 1988], where gapless chiral edge excitations can be solved exactly (figure 1.2). In general, there ν chiral modes along each edge and there is alway a pair of edges connecting two

---

[3]The magnetic Brillouin zone is topologically a torus. At each momentum $\mathbf{k}$ in the Brillouin zone, the valence band states span a vector space $V_{\mathbf{k}}$. The collection of these forms a vector bundle over the Brillouin zone. Vector bundles over a torus are stably classified by the first Chern number.
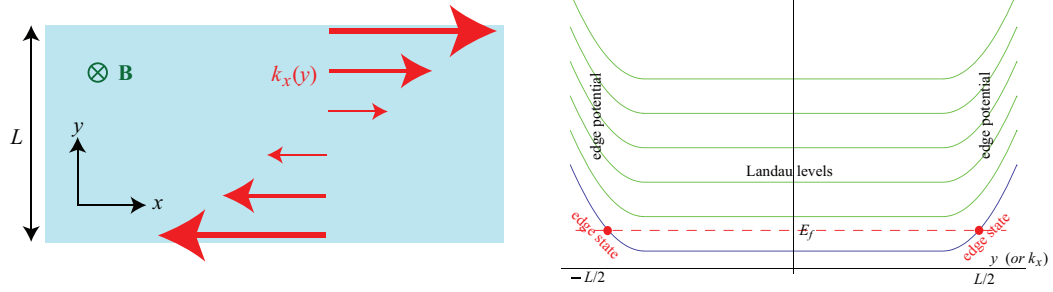
Figure 1.1: (A) Spatial configuration of momentum eigenstates in a Landau level; (B) Landau levels and band bending at edges.
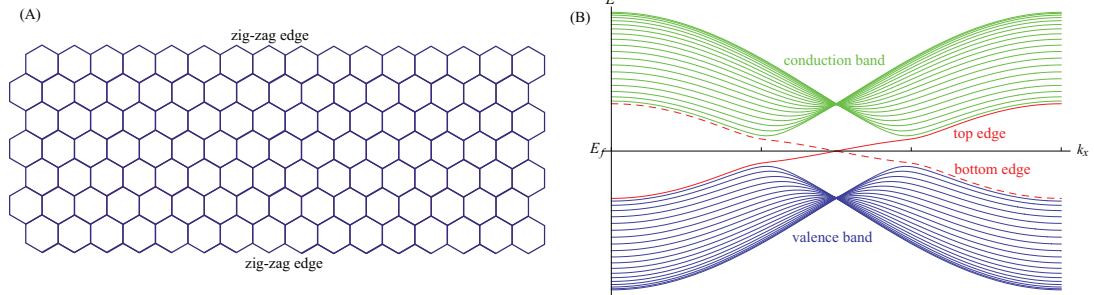
Figure 1.2: (A) A graphene slab terminating along the zig-zag edge; (B) chiral boundary excitations (solid red: right-moving mode along top edge, dotted red: left-moving mode along bottom edge).

charge reservoirs. This recovers the quantized Hall conductance eq.(1.1.1).

This brings us to the bulk-boundary correspondence that equates the TKNN integer ν, a topological property of the bulk insulating state, and the *chirality* of the gapless excitations along a single edge.

$$\nu = N_R - N_L = \text{number of right movers} - \text{number of left movers} \qquad (1.1.7)$$

It is a special case of the Atiyah-Patodi-Singer index theorem [Atiyah, Patodi and Singer, 1973]. The general proof of this involves equating the chirality of the edge mode and the *spectral flow* in a semi-infinite system, which can be expressed as the

11

winding number of a single particle Green's function [Volovik, 2003].

$$
\begin{aligned}
N_R - N_L &= \frac{1}{2} \left(\eta|_{k_x=-\infty} - \eta|_{k_x=+\infty} \right) & (1.1.8) \\
&= \frac{i}{2\pi} \oint d\omega \mathrm{Tr} \left(\hat{\mathcal{G}}(\omega, k_x) \right) \Big|_{k_x=-\infty}^{k_x=+\infty} & (1.1.9) \\
&= \frac{1}{24\pi^2} \oint d\omega \int d^2 k \mathrm{Tr} \left(\varepsilon_{\mu\nu\lambda} \mathcal{G}^{-1} \partial_\mu \mathcal{G} \mathcal{G}^{-1} \partial_\nu \mathcal{G} \mathcal{G}^{-1} \partial_\lambda \mathcal{G} \right) & (1.1.10)
\end{aligned}
$$

Eq.(1.1.10) coincides with the 1-loop term corresponding to parity anomaly in (2+1) dimensions [Alvarez-Gaumé, *et.al.*, 1985; Nakahara, 1990]. By substituting $\mathcal{G}(\omega, \mathbf{k}) = [\omega - H(\mathbf{k})]^{-1}$, it reduces to the TKNN integer in eq.(1.1.2).

1.1.2   Topological insulators in two and three dimensions

In two dimensions, time reversal symmetry guarantees the vanishing of the TKNN invariant since,

$$
\mathrm{Tr}(\mathcal{F}(-\mathbf{k})) = \mathrm{Tr}(\mathcal{F}(\mathbf{k}))^* = -\mathrm{Tr}(\mathcal{F}(\mathbf{k})) \tag{1.1.11}
$$

A new invariant is needed for the topological classification of time reversal symmetric band Hamiltonians [Kane and Mele, 2005b; Fu and Kane, 2006, 2007; Fukui and Hatsugai, 2007; Moore and Balents, 2007; Fukui, Fujiwara and Hatsugai, 2008; Qi, Hughes and Zhang, 2008; Roy, 2006a; Wang, Qi and Zhang, 2010].

In a spin-orbit coupled system, the canonical time reversal operator is $\Theta = i\sigma_y K$, where σ_y operates on spin and K is complex conjugation under the $|\uparrow\rangle, |\downarrow\rangle$ basis. The Kramer's theorem states that energy eigenstates must come in pairs $\{|\psi\rangle, \Theta|\psi\rangle\}$, or else Θ would square to $+1$ in the eigenspace. This implies at a time reversal invariant momentum (TRIM) where $\mathbf{k} = -\mathbf{k} + \mathbf{G}$, for $\mathbf{G}$ a reciprocal lattice

vector, Bloch states $|u_m(\mathbf{k})\rangle$ are doubly degenerate.

Assuming the bulk-boundary correspondence, topological invariants can be formulated by the boundary spectral behavior. For a d-dimensional bulk material, the boundary spectrum is defined on a $(d-1)$-dimensional projected surface Brillouin zone (fig.1.3), in which there are 2^{d-1} number of surface TRIM. A mid-gap Fermi
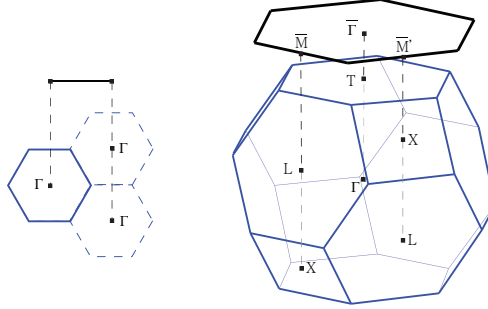
Figure 1.3: Projected surface Brillouin zone (black) and the bulk Brillouin zone (blue) of a 2D honeycomb lattice with a zig-zag edge and a 3D trigonal lattice terminated along the (111)-surface

energy ε_f cuts across the boundary spectrum to form a boundary Fermi surface in the projected surface Brillouin zone. Time reversal symmetry requires that the boundary Fermi surface must enclose a region in the surface Brillouin zone. The $\mathbb{Z}_2$-index is determined by whether the number of enclosed TRIM is even or odd.

$$\nu \equiv \text{number of enclosed TRIM in the boundary Fermi surface} \quad (\text{mod } 2) \quad (1.1.12)$$

Since the energy bands at the TRIM must form Kramer's doublets, any deformation of the surface bands or the Fermi energy can only alter the number ν by a multiple of 2. Therefore the $\mathbb{Z}_2$-index is robust against any perturbation that respects time reversal symmetry and maintains the bulk gap.[4]

---

[4]The $\mathbb{Z}_2$-index eq.(1.1.12) works only for the physical bulk dimensions $d = 2, 3$. For $d = 1$, the boundary spectrum is removable by surface potential. For $d = 4$, a $\mathbb{Z}$-index can be defined by the Chern invariant

$$\nu = \frac{i}{2\pi} \int_{\text{boundary Fermi surface}} \text{Tr}(\mathcal{F}) \in \mathbb{Z} \qquad (1.1.13)$$
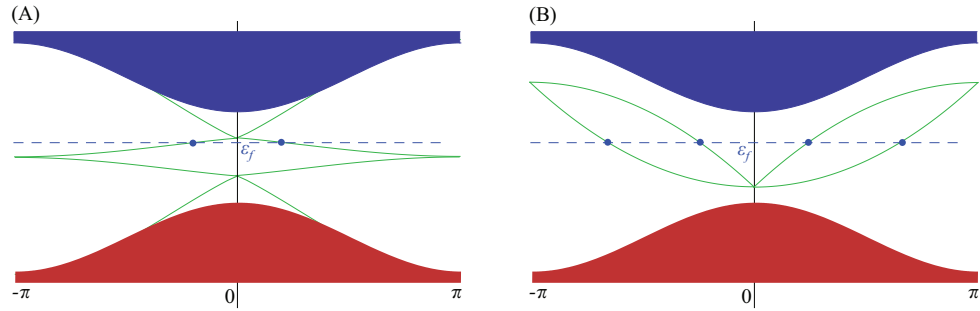
Figure 1.4: Edge spectrum of a 2D insulator. Fermi energy ε_f cuts Fermi points (blue dots) across the edge bands (green). TRIM are 0 and π. Degeneracies at the TRIM are protected due to Kramer's theorem. (A) Helical edge mode of a topological insulator. Fermi points enclose an odd number of TRIM, $\nu = 1$. (B) Non-helical edge mode of a trivial insulator. Fermi points enclose an even number of TRIM, $\nu = 2$.
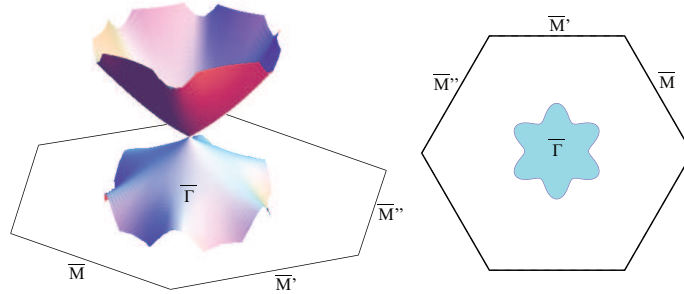
Figure 1.5: (Left) Surface Dirac cone spectrum of a 3D topological insulator. Degeneracy at the Dirac cone center is protected by Kramer's theorem. (Right) Fermi energy cuts a Fermi arc (blue loop) in the surface Brillouin zone. The TRIM $\bar{\Gamma}$ is enclosed and $\nu = 1$.

Perturbation and gauge insensitive bulk $\mathbb{Z}_2$-topological invariants can be constructed with the Bloch states $|u_i(\mathbf{k})\rangle$ of the bulk [Fu and Kane, 2006; Fu, Kane and Mele, 2007].

$$(-1)^\nu = \prod_{\text{TRIM } \mathbf{k}} \frac{\text{Pf}(w(\mathbf{k}))}{\sqrt{\det(w(\mathbf{k}))}}, \quad w_{ij}(\mathbf{k}) = \langle u_i(\mathbf{k})|\Theta u_j(-\mathbf{k})\rangle \qquad (1.1.14)$$

Here the product is taken over the TRIM at which the Pfaffian can be defined for the antisymmetric $w_{ij}(\mathbf{k})$. The sign of the square root is defined consistently so that $\sqrt{\det(w(\mathbf{k}))}$ lives on the same branch across the whole Brillouin zone.[5]

In two dimensions, the $\mathbb{Z}_2$-invariant eq.(1.1.14) can be shown to be identical to the integral invariant [Fu and Kane, 2006]

$$\nu \equiv \frac{i}{2\pi} \left[\int_{BZ_{1/2}} \text{Tr}(\mathcal{F}) - \oint_{\partial BZ_{1/2}} \text{Tr}(\mathcal{A}) \right] \pmod 2 \qquad (1.1.16)$$

where $BZ_{1/2}$ denotes half of the Brillouin zone and is chosen so that its boundary $\partial BZ_{1/2}$ is closed under time reversal, and the gauge dependent Berry connection $\mathcal{A}_{ij} = \langle u_i|du_j\rangle$ [Berry, 1984] is defined by a valence frame $|u_i(\mathbf{k})\rangle$ that satisfies a gauge constraint

$$w(\mathbf{k}) \equiv \text{constant}, \quad \mathbf{k} \in \partial BZ_{1/2}. \qquad (1.1.17)$$

In three dimensions, the $\mathbb{Z}_2$-invariant eq.(1.1.14) can be shown indirectly to be identical to the Chern-Simons invariant [Qi, Hughes and Zhang, 2008; Wang, Qi

---

where $\mathcal{F}$ is the Berry curvature [Berry, 1984] of boundary states. For $d \geq 5$, the number of enclosed TRIM must be even.

[5]In 2 and 3D, one can always choose a continuous valence frame $|u_i(\mathbf{k})\rangle$ over the entire Brillouin zone because the topological obstruction, the 1st Chern number, is killed by time reversal symmetry. Eq.(1.1.14) can be gauged away in 1D, and $\sqrt{\det(w(\mathbf{k}))}$ cannot be continuously defined for $d \geq 4$ when there is a non-vanishing higher Chern invariant. In fact, the $\mathbb{Z}$-invariant ν in 4D is defined by the second Chern number

$$\nu = \frac{1}{8\pi^2} \int_{BZ} \text{Tr}(\mathcal{F} \wedge \mathcal{F}) \qquad (1.1.15)$$

and Zhang, 2010]

$$\nu_0 = \frac{\theta}{\pi} \equiv \frac{1}{4\pi^2} \int_{BZ} \mathrm{Tr} \left(\mathcal{A} \wedge d\mathcal{A} + \frac{2}{3} \mathcal{A} \wedge \mathcal{A} \wedge \mathcal{A} \right) \quad (\mathrm{mod}\ 2) \qquad (1.1.18)$$

Both invariants (1.1.16, 1.1.18) are integer-valued because of time reversal symmetry, and any gauge transformation or perturbation can only alter them by integral multiples of 2.

Weak topological insulators The $\mathbb{Z}_2$-index ν_0 in eq.(1.1.18) (or equivalently eq.(1.1.14)) for a 3D insulator is called a *strong* topological invariant. It is inherited from the true 3D nature of the material as the Chern-Simons invariant involves a 3D integration. There are other types of invariants originating from layered structures, where 2D topological insulators are stacked along some normal direction. In momentum space, these lower dimensional invariants arise from the $\mathbb{Z}_2$-indices of 2D momentum planes closed under time reversal symmetry in the Brillouin zone. Since planes can be stacked in three directions, there are three *weak* indices $(\nu_1 \nu_2 \nu_3)$, which form a $\mathbb{Z}_2$-valued reciprocal lattice vector[6] [Fu, Kane and Mele, 2007; Fu and Kane, 2007]

$$\mathbf{G}_\nu = \nu_1 \mathbf{b}_1 + \nu_2 \mathbf{b}_2 + \nu_3 \mathbf{b}_3 \qquad (1.1.20)$$

---

[6]**Definition of weak indices** A primitive Bravais lattice vector $\mathbf{R} = m_1 \mathbf{a}_1 + m_2 \mathbf{a}_2 + m_3 \mathbf{a}_3$, for relatively prime m_1, m_2, m_3, defines a stratification of sub-tori T^2 in the Brillouin zone. A more sophisticated way to rephrase this is by identifying the Bravais lattice with the 2nd-homology $H_2(BZ)$, which contains classes of non-trivial 2-cocycles. The usual physical convention is to take the sub-torus $T^2_{\mathbf{R}}$ perpendicular to $\mathbf{R}$, which is closed under time reversal, and does not contain the origin Γ in the Brillouin zone. The weak index vector $\mathbf{G}_\nu$ is defined so that

$$\frac{1}{2\pi} \mathbf{G}_\nu \cdot \mathbf{R} = \frac{i}{2\pi} \left[\int_{(T^2_{\mathbf{R}})_{1/2}} \mathrm{Tr}(\mathcal{F}) - \oint_{\partial (T^2_{\mathbf{R}})_{1/2}} \mathrm{Tr}(\mathcal{A}) \right] \quad (\mathrm{mod}\ 2) \qquad (1.1.19)$$
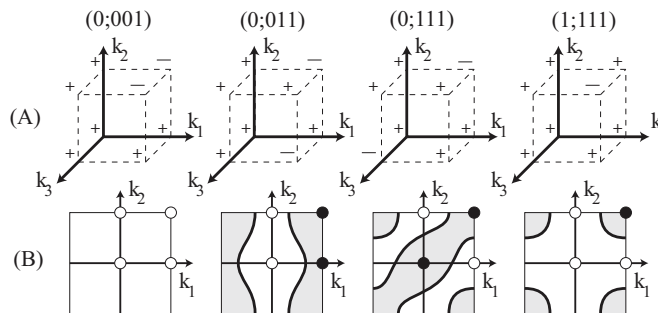
Figure 1.6: Diagrams of four phases with indices $(\nu_0; \nu_1\nu_2\nu_3)$. (A) Values of $\delta(\Gamma_i) = \mathrm{Pf}(w(\Gamma_i))/\sqrt{\det(w(\Gamma_i))}$ at the 8 TRIM. (B) Surface band Fermi arcs along the (001) surface.

1.1.3 Inversion symmetry

Inversion symmetry helps in determining the topological phase of insulators [Fu and Kane, 2007]. It greatly reduces the computation effort of calculating the $\mathbb{Z}_2$-invariants $(\nu_0; \nu_1\nu_2\nu_3)$ by only requiring the knowledge of parity eigenvalues at the inversion symmetric momenta (a.k.a. TRIM). This avoids solving a continuous set of Bloch wave functions across the Brillouin zone.

Inversion symmetry is represented by a non-local unitary operator P. By a gauge transformation if necessary, we may assume $P^2 = 1$. In an insulator, the parity of an inversion symmetric momentum Γ_i is given by a product of inversion eigenvalues of valence states $|u_i(\Gamma_i)\rangle$. The simplest topological non-trivial inversion symmetric band theory is given by the Su-Schrieffer-Heeger model of polyacetylene [1980]. Depending on the relative strength of the bonds and position of the inversion
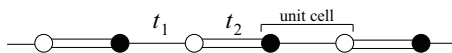
Figure 1.7: Su-Schrieffer-Heeger model: $H(k) = (t_1 + t_2 \cos k)\sigma_x + t_2 \sin k\sigma_y$ and $P = \sigma_x$ with inversion center at the weak bond or $P' = e^{ik}\sigma_x$ with inversion center at the strong bond.

center, the parity at $k = 0, \pi$ can be the same or opposite.

Inversion commutes with time reversal, and therefore the inversion eigenvalues are the same in a Kramer's doublet $\{|u_{2m-1}(\Gamma_i)\rangle, |u_{2m}(\Gamma_i)\rangle = \Theta|u_{2m-1}(\Gamma_i)\rangle\}$. The parity invariant is defined by taking the product over half the valence states

$$\delta(\Gamma_i) = \prod_{E_{2m}<E_f} \xi_{2m}(\Gamma_i), \quad \xi_{2m}(\Gamma_i) = \langle u_{2m}(\Gamma_i)|Pu_{2m}(\Gamma_i)\rangle \qquad (1.1.21)$$

The time reversal $\mathbb{Z}_2$-invariant can be deduced from the parity eigenvalues at the inversion fixed points.

$$(-1)^\nu = \prod_{\text{TRIM } \Gamma_i} \delta(\Gamma_i) \qquad (1.1.22)$$

Chapter 2

Surface spectrum of bismuth antimony alloy

Pure Bi or Sb are semimetals [Golin, 1968a,b; Liu and Allen, 1995; Lenoir *et.al.*, 1996]. They have distinct conduction and valence bands, being separated by a direct gap but overlap in energies at different momenta. An insulator phase was discovered for the $\text{Bi}_{1-x}\text{Sb}_x$, for $0.07 \lesssim x \lesssim 0.22$ [Lenoir *et.al.*, 1996]. A metallic phase transition occurs at $x \approx 0.04$ where bands invert at the three L points of the Brillouin zone, and this changes the topology. The topological phase can be read off from the parity of the inversion fixed points of the materials, and it was predicted that the insulating alloy $\text{Bi}_{1-x}\text{Sb}_x$ was a strong topological insulator [Fu and Kane, 2007]. Here we provide numerical calculation on surface excitations that support the prediction [Teo, Fu and Kane, 2008].

The electronic surface band spectrum of bismuth antimony alloy $\text{Bi}_{1-x}\text{Sb}_x$ is

	$\delta(\Gamma)$	$\delta(L)$	$\delta(T)$	$\delta(X)$	$(\nu_0; \nu_1\nu_2\nu_3)$
Bi	-1	-1	-1	-1	(0;000)
Sb	-1	1	-1	-1	(1;111)
$Bi_{1-x}Sb_x$	-1	1	-1	-1	(1;111)

Table 2.1: Parity invariants $\delta(\Gamma_i)$ at the TRIM (see fig.1.3 for the Brillouin zone of the trigonal lattice) and $\mathbb{Z}_2$-indices $(\nu_0; \nu_1\nu_2\nu_3)$ for bismuth, antimony [Liu and Allen, 1995], and $Bi_{1-x}Sb_x$ [Fu and Kane, 2007].

studied using a tight binding model proposed by Liu and Allen [1995], and is solved by us numerically in a semi-infinite geometry. The main conclusions include the following.

1. **Topological structure of surface bands** The alloy is shown to be a strong topological insulator by the Dirac spectrum along the (111) trigonal surface.

2. **Inversion symmetry and Fermi surfaces** The bulk parity invariants $\delta(\Gamma_i)$ of the 8 TRIM determines topologically how the 4 surface TRIM are enclosed by the Fermi arcs along an arbitrary surface.

3. **Mirror Chirality** A new topological invariant known as *mirror Chern number* is introduced using mirror symmetry of the material, and a sign inconsistency is discovered when compared with an earlier pseudo-potential results and first principle calculation on pure Bi. This is due to an artifact of the tight binding model that focuses on the energy spectrum rather than the topology of states.
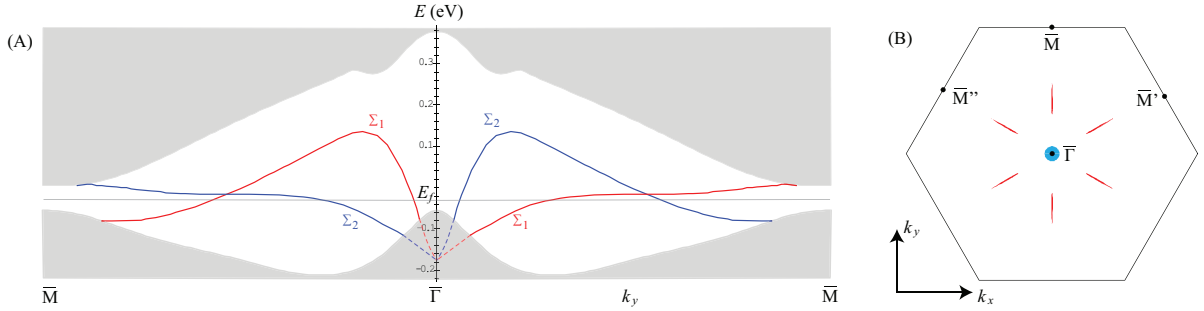
Figure 2.1: (A) Tight-binding generated surface band spectrum of $Bi_{1-x}Sb_x$, for $x = 0.08$, along the bisectrix $\bar{\Gamma}\bar{M}$ direction. Surface bands Σ_1, Σ_2 are labeled by opposite mirror eigenvalues, and have opposite chirality: Σ_1 is forward moving and Σ_2 is backward moving. (B) Electron pocket (blue) and hole pockets (red) on the surface Brillouin zone.

2.1 Topological structure of surface bands

The surface spectrum predicted by the tight binding model shows 7 Dirac cones on the surface Brillouin zone. One of them sits at the TRIM $\bar{\Gamma}$, and the other 6 are located midway along the bisectrix directions related to each other by 3-fold and time reversal symmetries. The odd number of surface Dirac conse $Bi_{1-x}Sb_x$ suggests it is a strong topological insulator. The Dirac cone at $\bar{\Gamma}$ is protected by Kramer's theorem and is submersed in the valence band (fig.2.1(A)). This forms an electron pocket surrounding $\bar{\Gamma}$ in the surface Brillouin zone (fig.2.1(B)). The surface band crossings (fig2.1(A)) above Fermi energy midway along $\bar{\Gamma}\bar{M}$ is protected by mirror symmetry, and these forms the 6 Dirac cones that surrounds 6 hole pockets on the surface Brillouin zone.

The surface states are not spin eigenstates, but they have a non-zero spin expectation value as a function of momentum, forming a spin texture around the Dirac cones (fig.2.2). Along the mirror symmetric bisectrix $\bar{\Gamma}\bar{M}$ axis (the y-axis), the spin expectation must be proportional to the perpendicular binary x-axis $\langle \vec{S} \rangle \propto$
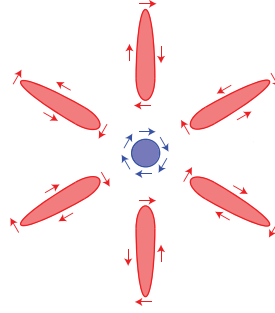
Figure 2.2: Schematic of spin texture around the electron and hole pockets (not to scale).

$i\langle M\rangle \hat{x} \propto -(+)\hat{x}$ for mirror eigen-sectors $\Sigma_{1(2)}$. This is because the mirror operator $M = P\exp(-i\pi(L_x + S_x)/\hbar)$ anticommutes with the other spin operators $M^\dagger S_{y/z}M = -S_{y/z}$. Spins rotate by 2π around the electron and hole pockets, and this confirms the Dirac structure by the π Berry phase.

Shortly after the tight binding derivation of surface spectrum, an ARPES experiment was carried out on BiSb alloy and showed the strong topological signature of the insulator [Hsieh *et.al.*, 2008, 2009].
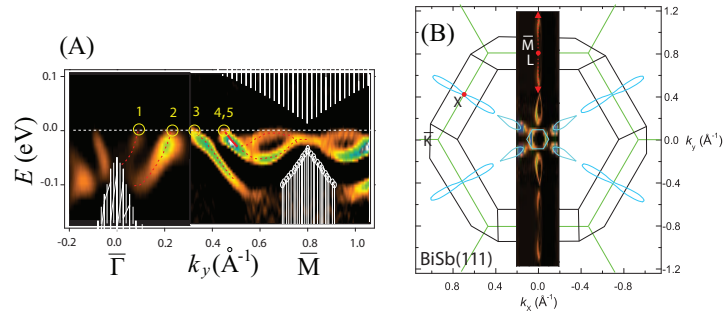
Figure 2.3: Surface spectrum of BiSb alloy from ARPES experiment [Hsieh *et.al.*, 2008].

2.2 Inversion symmetry and Fermi surfaces

The configuration of electron or hole pockets at the surface TRIM is set by inversion symmetry. The bulk parity invariants eq.(1.1.21) determines the *surface fermion parity* at the surface TRIM $\bar{\Lambda}$

$$\pi(\bar{\Lambda}) = (-1)^{N(\bar{\Lambda})} = (-1)^n \delta(\Gamma_1)\delta(\Gamma_2) \tag{2.2.1}$$

where the bulk TRIM Γ_1, Γ_2 project to the surface TRIM $\bar{\Lambda}$, and n is the number of bulk Kramer's degenerate valence bands. Here $N(\Lambda)$ counts the surface fermion number[1] with momentum $\bar{\Lambda}$ below the Fermi energy. And therefore the surface fermion parity $\pi(\bar{\Lambda})$ tells us the eveness or oddness of the number of electron or hole pockets surrounding $\bar{\Lambda}$ in the surface Brillouin zone.

The bulk parity invariants $\delta(\Gamma_i)$ are defined by an inversion center lying on the crystal boundary surface. For example, a dangling bound state appears at the end of a Su-Schrieffer-Heeger model (fig.1.7) when it is terminated at a strong bond. The boundary fermion parity is therefore -1, which agrees (2.2.1).

The surface fermion parity gives extra knowledge that cannot be extracted just from the topological phase of the insulator. For example, pure Bi is in the trivial phase with indices $(0; 000)$. But its bulk parity invariants predict all surface TRIM of the (111) surface are enclosed by electron or hole pockets. This ensures the existence of surface bands if there is no surface reconstruction and no net charge

---

[1]This number is an integer because of inversion and time reversal symmetry. Consider a periodic crystal of finite number of unit cells. The surface states at some inversion symmetric cut plane of the crystal are shared equally by both sides of the plane due to inversion symmetry. Since bands are doubly degenerate due to time reversal, N is an integer.

Surface	$\bar{\Lambda} = (\Gamma_1\Gamma_2)$	$\pi_{\text{BiSb}}(\bar{\Lambda})$	$\pi_{\text{Bi}}(\bar{\Lambda})$
(111)	$\bar{\Gamma} = (\Gamma T)$	-1	-1
	$3\bar{M} = (LX)$	$+1$	-1
(111)′	$\bar{\Gamma} = (\Gamma T)$	$+1$	$+1$
	$3\bar{M} = (LX)$	-1	$+1$
(110)	$\bar{\Gamma} = (\Gamma X)$	-1	-1
	$\bar{X}_1 = (LL)$	-1	-1
	$\bar{X}_2 = (LT)$	$+1$	-1
	$\bar{M} = (XX)$	-1	-1
(100)	$\bar{\Gamma} = (\Gamma L)$	-1	$+1$
	$\bar{M} = (TX)$	$+1$	$+1$
	$2\bar{M}' = (LX)$	-1	$+1$
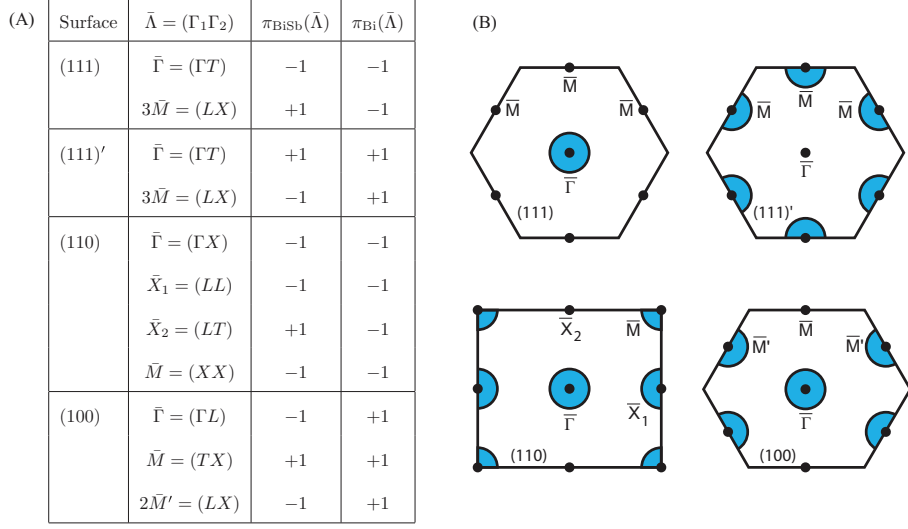
Figure 2.4: (A) Surface fermion parities of BiSb alloy or pure Bi cleaving at 4 boundary surfaces. (111)′ is the energetically unfavored cleavage along the (111) direction, where strong bonds are cut in between a bilayer. (B) Schematic showing which TRIM are enclosed by an odd number of electron or hole pockets for the 4 boundary surfaces of BiSb alloy.

accumulated on the surface, although the surface bands may not connect the bulk conduction and valence bands and they are removable by surface potential.

2.3   Mirror chirality

The crossing in figure 2.1(A) midway between $\bar{\Gamma}\bar{M}$ and the 6 Dirac hole pockets in fig.2.1(B) are protected from Mirror symmetry. The trigonal lattice of $\text{Bi}_{1-x}\text{Sb}_x$ has mirror planes normal to the binary x-axis. Hence the bisectrix-trigonal yz-plane in the Brillouin zone is mirror invariant. Bloch states along the $k_y k_z$-plane can be labeled by their mirror eigenvalues $M = P\exp(-i\pi(L_x + S_x)/\hbar) = \pm i$, each being time reversal of the other since $[\Theta, M] = 0$.

Analogous to spin Chern number [Sheng *et.al.*, 2006; Fu and Kane, 2006], a *mirror Chern number* $n_{\mathcal{M}}$ can be defined by one of the mirror eigensector, say
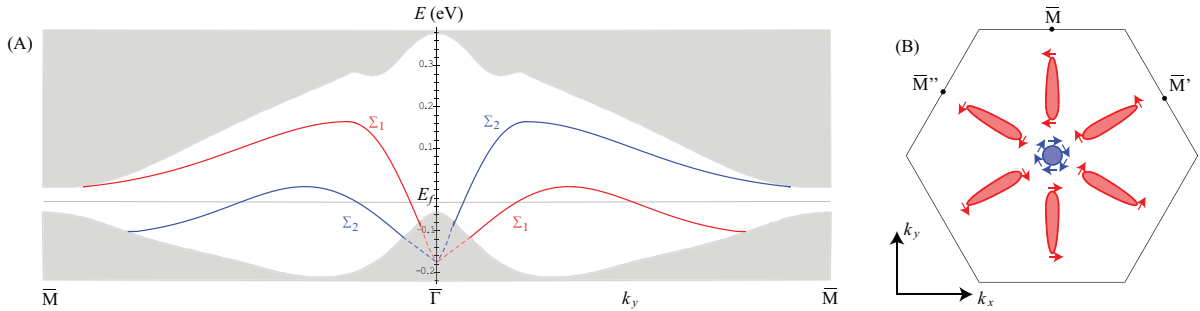
Figure 2.5: Schematic of a more realistic surface band spectrum with the correct mirror Chern number. (A) The chirality of the mirror bands are reversed, and there is no midway crossing in between $\bar{\Gamma}\bar{M}$. (B) There is only one Dirac cone sitting at $\bar{\Gamma}$. The Berry phase around the hole pockets (red) is zero.

$M = +i$. (The Chern number of the other sector $M = -i$ is negative of that of $+i$ due to time reversal.) The eveness and oddness of the mirror Chern number determines the strong topological phase of the 3-fold symmetric insulator. Indeed, this is shown by the chirality of the mirror surface modes in fig.2.1(A) so that

$$n_{\mathcal{M}} = +1 \qquad (2.3.1)$$

This however contradicts with a pseudopotential result [Golin, 1968a] and 1st principle calculations on pure Bi [Hirahara $et.al.$, 2007; Koroteev $et.al.$, 2008] that suggests a different sign of the mirror Chern number. As confirmed in the later ARPES observation on BiSb alloy [Hsieh $et.al.$, 2008], the mirror Chern number should instead be opposite $n_{\mathcal{M}} = -1$. Although this does not affect the topological phase of the material, the number of surface Dirac cones reduces as the 6 hole pockets are no longer protected from the π Berry phase.

Chapter 3

Point contact in a quantum spin

Hall insulator

Figure 3.1: Spin polarized helical edge modes of a quantum spin Hall insulator (QSHI).
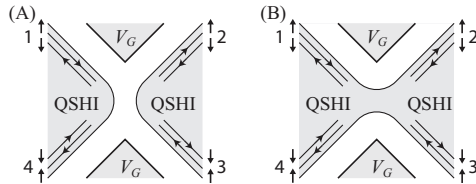
Figure 3.2: A quantum point contact in a QSHI controlled by a gate voltage V_G that closes in (A) for $V_G < V_G^*$ or opens in (B) for $V_G > V_G^*$. 1,2,3,4 mark the four terminals that connect to the point contact.

The edge of a quantum spin Hall insulator (QSHI) carries spin polarized helical modes [Kane and Mele, 2005a,b; Wu, Bernevig and Zhang, 2006; Xu and Moore, 2006; König *et.al.*, 2007], being essentially half of a metallic 1D electronic mode with spin. With electron-electron interaction, the edge mode is described by a *helical Luttinger liquid* theory [Hou *et.al.*, 2009; Ström and Johannesson, 2009], which is

like a spinless Luttinger liquid except localization is impossible as backscattering is forbidden by time reversal. A pair of edge modes are described by a *spinful Luttinger liquid* (SLL) theory [Kane and Fisher, 1992a; Furusaki and Nagaosa, 1993], in which the charge and spin Luttinger parameters are reciprocal to each other $g = g_\rho = 1/g_\sigma$.[1] A powerful tool for probing edge transport behavior is by using a quantum point contact (fig.3.2), where 4 terminal conductance can be measured as a function of temperature T and the gate voltage V_G that control the opening and closing of the contact.

With time reversal and spin rotational symmetry, the 4 terminal conductance matrix G_{ij} (defined by $I_i = G_{ij}V_j$, $i = 1, \ldots, 4$) can be re-expressed by 3 conductances G_{XX}, G_{YY} and $G_{XY} = G_{YX}$.

$$\begin{pmatrix} I_X \\ I_Y \\ I_Z \end{pmatrix} = \begin{pmatrix} G_{XX} & G_{XY} & 0 \\ G_{YX} & G_{YY} & 0 \\ 0 & 0 & 2e^2/h \end{pmatrix} \begin{pmatrix} V_X \\ V_Y \\ V_Z \end{pmatrix} \tag{3.0.1}$$

where the directional currents and potential differences are defined according to figure 3.3. The diagonal conductances G_{XX} and G_{YY} are simply the 2 horizontal
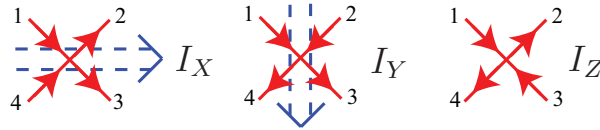
Figure 3.3: Currents I_X, I_Y, I_Z across the 4 terminal point contact.

or vertical terminal conductances, and they are identical

$$G^*_{XX} = G^*_{YY} = G^* \tag{3.0.2}$$

---

[1] The Luttinger parameter indicates the strength of electrons interaction, $g = 1$ for no interaction, $g < 1$ ($g > 1$) for repulsive (attractive) interaction.

at the transition V_G^* between pinching off and openning the point contact. The skew conductance G_{XY} is killed by mirror symmetry about the horizontal or vertical axis,[2] but when present, has different temperature exponent dependence (even for weak interaction $g \sim 1$) from the diagonal conductances [Teo and Kane, 2009]. This feature is unique in a spin-filtered helical mode, and thus provides a diagnostic for the topological nature of QSHI.

We are interested in critical conductance behavior [Teo and Kane, 2009]. The issue is to (i) identify the stable transport phases associated with the point contact, and (ii) investigate the crossover of conductance in between the stable phases. The foundation on studying these is the SLL theory of the gapless edge helical modes. The point contact is described by quantum Brownian motions [Yi and Kane, 1998], sine-Gordon type potentials of the bosonic charge and spin fields, that describe backscattering or tunneling of charge or spin across the contact. A renormalization group (RG) calculation then identifies the phases by determining the relevance of the fixed points of the theory.

3.1   Stable transport phases

With spin rotation symmetry, there are 4 stable phases listed in table 3.1. They describe the transport behavior of charge and spin across the QSHI in the X-direction. For example when the point contact is completely open (fig.3.2(A)), it is in the charge conducting/spin conducting (CC) phase; or when the point contact is

---

[2]Mirror symmetry will in general be absence in a real life point contact. But an *emergent* Mirror symmetry arises in the low energy effective theory of the stable fixed points.

phases	II	CC	IC	CI
charge	insulating	conducting	insulating	conducting
spin	insulating	conducting	conducting	insulating
g	$g < 1/2$	$1/2 < g < 2$	$1/2 < g < 2$	$g > 2$
G_{XX}	0	$2e^2/h$	0	$2e^2/h$
G_{YY}	$2e^2/h$	0	0	$2e^2/h$
G_{XY}	0	0	0	0

Table 3.1: Stable phases of QSHI point contact (with spin rotation symmetry), the stability region for the Luttinger parameter g, and the zero temperature directional conductances across the point contact.
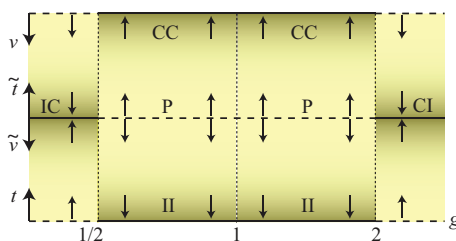
Figure 3.4: Phase diagram for a point contact in a QSHI as a function of the Luttinger parameter g. The arrows indicate the stability of the CC, II, CI and IC phases, as well as the critical fixed point P. $v, \tilde{v}, t, \tilde{t}$ are backscattering and tunneling terms of the point contact.

completely pinched-off, it is in the charge insulating/spin insulating (II) phase.

The II and CC phases are both stable for weak interaction, where the Luttinger parameter lies between $1/2 < g < 2$. This means that in between the pinch-off to the open limit, the point contact must go through an intermediate phase that is characterized by an intermediate fixed point[3] [Kane and Fisher, 1992a; Furusaki and Nagaosa, 1993]. The critical behavior of this intermediate fixed point controls the crossover conductance, and it is solvable in the three limits: (i) $g = 1 - \epsilon$, (ii) $g = 1/2 + \epsilon$ and (iii) $g = 1/\sqrt{3}$ (results of $g > 1$ can be extracted from that of $g < 1$ by a duality).

---

[3]This feature is novel and is absent in a typical Luttinger liquid [Kane and Fisher, 1992a,b; Furusaki and Nagaosa, 1993] or fractional quantum Hall point contact [Moon *et.al.*, 1993; Milliken *et.al.*, 1996].
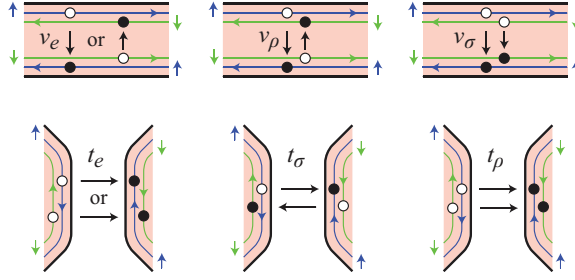
Figure 3.5: Least irrelevant electronic backscattering and tunneling terms of a point contact in the CC (top) and II (bottom) phase. (Top left to right) v_e backscatters a charge and a spin; v_ρ backscatters 2 charges no spin; v_σ backscatters no charge 2 spins. (Bottom left to right) t_e tunnels a charge and a spin; t_σ tunnels no charge 2 spins; t_ρ tunnels 2 charges no spin.

phases	II		CC		IC	CI
δG_{XX}	$\sim$	$\begin{cases} T^{g+g^{-1}-2}, & g < \sqrt{3} \\ T^{4/g-2}, & g > \sqrt{3} \end{cases}$	$\sim$	$\begin{cases} -T^{g+g^{-1}-2}, & g > 1/\sqrt{3} \\ -T^{4g-2}, & g < 1/\sqrt{3} \end{cases}$	$\sim T^{g^{-1}-2}$	$\sim T^{g-2}$
δG_{YY}	$\sim \begin{cases} -T^{g+g^{-1}-2}, & g > 1/\sqrt{3} \\ -T^{4g-2}, & g < 1/\sqrt{3} \end{cases}$		$\sim \begin{cases} T^{g+g^{-1}-2}, & g < \sqrt{3} \\ T^{4/g-2}, & g > \sqrt{3} \end{cases}$		$\sim T^{g^{-1}-2}$	$\sim T^{g-2}$
δG_{XY}	$\sim T^{g+g^{-1}}$		$\sim T^{g+g^{-1}}$		$\sim T^{2g^{-1}-2}$	$\sim T^{2g-2}$

Table 3.2: Low temperature dependence of conductances at the stable phases.

The low energy behavior of the point contact can be described by electronic backscattering and tunneling terms in the SLL theory (such as in fig.3.5). They are all irrelevant perturbations of the stable fixed points of the point contact, but contribute to conductances in finite temperature. For example, the backscattering potential v_e gives a perturbation proportional to v_e^2 in the diagaonal conductances G_{XX} and G_{YY} (from the Fermi golden rule), and this coupling constant flows according to the energy scale of the system. The small temperature dependence at the stable phases can therefore be read off from the scaling dimensions $\Delta(v)$ (or $\Delta(t)$) of the least irrelevant backscattering (or tunneling) operators and the RG equation

$$\frac{dv}{dl} = (1 - \Delta(v))v \tag{3.1.1}$$

30
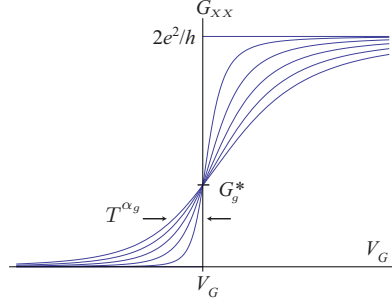
Figure 3.6: Conductance G_{XX} as a function of gate voltage V_G for different temperatures. As the temperature is lowered, the pinch-off curve sharpens up with a width T^{α_g}. The curves cross at a critical conductance G_g^*, and the shape of the curve has the universal scaling form (3.2.3).

Finally, if spin rotation symmetry is broken (by spin orbit interaction or multiple spin flip processes), the system flows to a time reversal breaking phase at low energy for strong interaction $g < 1/2$. A 2 spin flip term becomes relevant for $g < 1/2$ and gaps out the the helical modes. The two edges of the QSHI becomes insulating so that the entire 4 terminal conductance G_{ij} matrix vanishes at zero temperature.

3.2 Critical behavior of conductance

Critical behavior of the crossover conductance $G_{XX}(V_G, T)$ (plotted in fig.3.6) between the pinch-off and open limit of the point contact is controlled by the intemediate fixed point P midway between the II and CC phase (fig.3.4). The temperature and gate voltage dependence of the conductance collapse onto a single universal scaling function $\mathcal{G}_g$.

$$\lim_{\Delta V_G, T \to 0} G_{XX}(V_G, T) = \frac{2e^2}{h} \mathcal{G}_g \left(c \frac{\Delta V_G}{T^{\alpha_g}} \right) \tag{3.2.3}$$
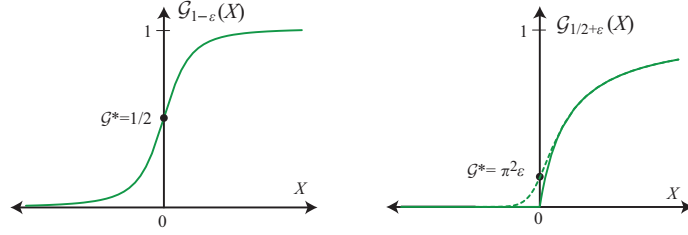
Figure 3.7: Universal crossover scaling functions $\mathcal{G}_g(X)$ for $g = 1 - \epsilon$ and $g = 1/2 + \epsilon$. On the right, dotted line is for $g = 1/2 + \epsilon$ and solid line is at $g = 1/2$. Graphs are plotted from the exact results

$$\mathcal{G}_{1-\epsilon}(X) = \frac{1}{2}\left(1 + \frac{X}{\sqrt{1 + X^2}}\right), \quad \mathcal{G}_{1/2+\epsilon}(X) = \frac{X}{X + 1 - e^{-X/(\pi^2\epsilon)}} \tag{3.2.1}$$
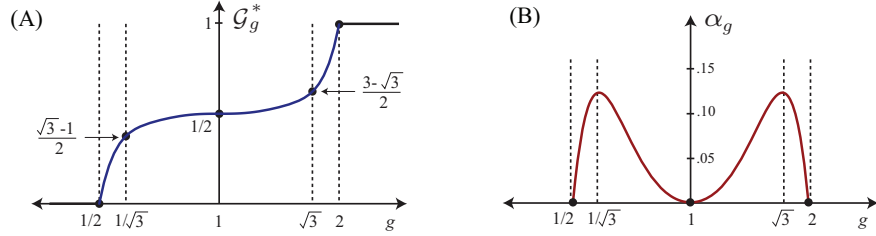
Figure 3.8: (A) Critical conductance (in units of $2e^2/h$) at critical gate voltage V_G^* and (B) critical exponent α_g as functions of g. The curves are polynomial fits, which incorporate the exact results at $g = 1 - \varepsilon$, $g = 1/2 + \varepsilon$ and $g = 1/\sqrt{3}$.

$$\mathcal{G}_g^* = \begin{cases} 1/2 + O(\epsilon^3) & \text{for } g = 1 - \epsilon \\ (\sqrt{3} - 1)/2 & \text{for } g = 1/\sqrt{3} \\ \pi^2\epsilon & \text{for } g = 1/2 + \epsilon \end{cases}, \quad \alpha_g = \begin{cases} \epsilon^2/2 & \text{for } g = 1 - \epsilon \\ 4\epsilon & \text{for } g = 1/2 + \epsilon \end{cases} \tag{3.2.2}$$

where $\Delta V_G = V_G - V_G^*$ and c is some non-universal constant.[4] The scaling function is symmetric about $g = 1$ by

$$\mathcal{G}_{1/g}(X) = 1 - \mathcal{G}_g(-X) \qquad (3.2.4)$$

due to the duality of the II and CC phase. This allows us to consider only repulsive interactions $g \leq 1$.

The universal scaling function $\mathcal{G}_g(X)$ is determined by the intermediate fixed point P in between the stable II and CC phases. It is solvable at the three limits: (1) $g = 1 - \epsilon$, (2) $g = 1/2 + \epsilon$ and (3) $g = 1/\sqrt{3}$. Perturbative RG analysis predicts the scaling behavior eq.(3.2.1) and the critical conductances $\mathcal{G}_g^* = \mathcal{G}_g(0)$ in eq.(3.2.2). The exact scaling behavior in the two solvable limits are plotted in fig.3.2. The critical conductance is fitted as a function of g using a minimal degree polynomial and is plotted in fig.3.7.

1. $\boxed{g = 1 - \epsilon}$ At the non-interacting limit, the 4 terminal conductances G_{ij} can be expressed in form of the scattering matrix

$$G_{ij} = \frac{e^2}{h}(\delta - |S_{ij}|^2), \quad |S_{ij}|^2 = \begin{pmatrix} 0 & \mathcal{T} & \mathcal{F} & \mathcal{R} \\ \mathcal{T} & 0 & \mathcal{R} & \mathcal{F} \\ \mathcal{F} & \mathcal{R} & 0 & \mathcal{T} \\ \mathcal{R} & \mathcal{F} & \mathcal{T} & 0 \end{pmatrix} \qquad (3.2.5)$$

where the S-matrix relates incomming and outgoing modes at leads i, j. A perturbative analysis up to second order in interaction $\epsilon = 1 - g$ gives the RG

---

[4]The other diagonal conductance is given by $G_{YY}(V_G, T) = (2e^2/h)\mathcal{G}_g(-X)$ for $X = cV_G/T^{\alpha_g}$.

equations for the S-matrix parameters and predicts the critical behavior of the crossover conductance eq.(3.2.1) at $g = 1 - \epsilon$.
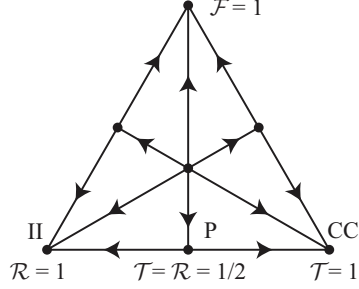
Figure 3.9: RG flow of the S-matrix parameters for weak interaction $g = 1 - \epsilon$. The point contact lies on the bottom line when there is spin symmetry, where the stable fixed points II and CC are separated by an intermediate fixed point P.

2. $\boxed{g = 1/2 + \epsilon}$ The IC $\leftrightarrow$ CC/II phase transition occurs at $g = 1/2$. The most relevant perturbations of the IC fixed point are $\tilde{v}_\sigma$, the backscattering of a spin, and $\tilde{t}_\rho$, the tunneling of a charge across the point contact. They become relevant when $g > 1/2$ and the system flows to the CC or II fixed point at low energy. A perturbative analysis up to second order in the perturbations gives the RG equation that locate the intermediate fixed point P at $\epsilon = g - 1/2$, and the Kubo formula predicts the conductance around the P

$$\mathcal{G}_{1/2+\epsilon}(X) = \frac{X}{1 - e^{-X/(\pi^2 \epsilon)}}, \quad \text{for small } X = \Delta V_G / T^{4\epsilon} \tag{3.2.6}$$

For large X, the point contact will lie on the horizontal axis of fig.3.10, where the entire IC to CC crossover can be derived by a fermionization procedure. The scaling function is

$$\mathcal{G}_{1/2+\epsilon}(X) = \theta(X)\frac{X}{X+1}, \quad \text{for large } X = \Delta V_G / T^{4\epsilon} \tag{3.2.7}$$

34

Eq.(3.2.6) and (3.2.7) combines into (3.2.1) at $g = 1/2 + \epsilon$.
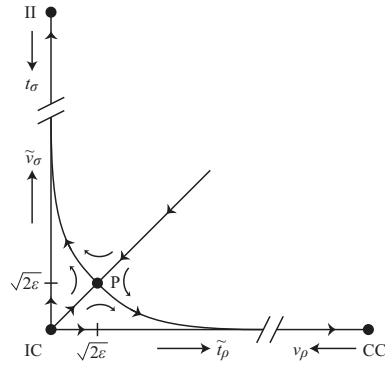
Figure 3.10: RG flow of for $g = 1/2 + \epsilon$.

3. $\boxed{g = 1/\sqrt{3}}$ The QBM model at $g = 1/\sqrt{3}$ has a triangular lattice symmetry, which implies the intermediate fixed point is self-dual under the II $\leftrightarrow$ CC duality. The critical conductance $\mathcal{G}^*_{1/\sqrt{3}}$ can then be solved self-consistently.

Chapter 4

Majorana fermions in three dimensions

Majorana fermion (or real fermion) is a fermion that is its own anti-particle [Majorana, 1937]. Neutrino could theoretically be an elementary particle in the standard model that behaves as a Majorana fermion, but its nature is still not completely clear as the observation of neutrinoless double beta decay has not been settled [Aalseth *et.al.*, 2002; Avignone, Elliott and Engel, 2008]. Another approach of studying Majorana fermions is through emergent quasiparticle excitations in condensed matter systems [Wilczek, 2009; Franz, 2010; Service, 2011]. Current promising candidates include half vortex excitation on a chiral p-wave superconducductor [Caroli, de Gennes and Matricon, 1964; Volovik, 1999; Read and Green, 2000; Kitaev, 2001; Ivanov, 2001; Stern, von Oppen and Mariani, 2004; Das Sarma, Nayak and Tewari, 2006; Stone and Chung, 2006; Nayak *et.al.*, 2008], surface excitation of the superfluid ^{3}He-B [Kopnin and Salomaa, 1991; Volovik, 2003; Chung and Zhang, 2009],

quasihole excitation on a non-Abelian 5/2 fractional quantum Hall fluid [Moore and Read, 1991; Greiter, Wen and Wilczek, 1992; Read and Green, 2000], and the recently proposed topological insulator - superconductor heterostructures [Fu and Kane, 2008] and strongly spin-orbit coupled semiconductor - superconductor heterostructures [Sau, *et.al.*, 2010; Alicea, 2010].

Majorana excitations have the desirable property of being robust against accidental measurements [Kitaev, 2003]. Unlike a Dirac fermion (or complex fermion), a pair of Majorana fermions is required to define a two level system. A quantum state cannot be measured unless the pair of Majorana fermions is brought together in space and allowed to *fuse* [Kitaev, 2006]. This means that quantum information is stored coherently by spatially separated Majorana excitations. Unitary operations can be constructed by manipulating the physical locations of the Majorana fermions (or by non trivial cycles in the configuration space). In $(2+1)$D, these are known to include the non-Abelian Ising exchange statistics [Moore and Read, 1991; Nayak and Wilczek, 1996; Ivanov, 2001]. These are the essential ingredients for constructing a *topological quantum computer*[1] [Das Sarma, Freedman and Nayak; Nayak *et.al.*, 2008].

The non-local nature of quantum states makes Majorana fermions hard to detect. There are multiple proposals for experimental observations in various electronic systems using techniques in interferometry or point contact [Das Sarma, Freedman and Nayak; Stern and Halperin, 2006; Akhmerov, Nilsson and Beenakker, 2009; Fu and Kane, 2009b], although indisputable evidence of the exotic excitation is still not

---

[1]$(2 + 1)$D Ising anyons are still not sufficient to generate enough unitary operations for a universal quantum computer [Freedman, Larsen and Wang, 2002].

yet available. The issue of detection will not be addressed in the thesis. Instead, the main focus will be on the physical nature of Majorana fermions as an excitation in three dimensions [Teo and Kane, 2010a].

1. What is the topological criterion for having a Majorana fermion?

2. What are the topologically protected unitary operations on a collection of Majorana fermions in 3D?

This section will be organized in the following way. It will begin by reviewing the structure of Majorana fermions and their origins in some previous condensed matter systems. More elaborate description will be given on topological insulator - superconductor heterostructures, which is one of the prime subjects in the thesis. A $\mathbb{Z}_2$ topological index will be introduced in the Bogoliubov de Gennes (BdG) description of the system, and it will serve as a mathematical indicator for the existence of a Majorana zero mode. The configuration space of a collection of spatially separated Majorana excitations will then be described with the help of an effective minimal 8 band model. It will be showed that Majorana fermions are interconnected through a 3-component vector field in real space, and its non-local nature opens the possibilty of topologically non-trivial exchange and braidless operations of Majorana fermions in three dimensions.

4.1 Majorana fermions in condensed matter systems

Majorana fermion operators are hermitian operators $\gamma_j = \gamma_j^\dagger$, that obey the canonical anticommutation relation for fermions

$$\{\gamma_i, \gamma_j\} = \gamma_i\gamma_j + \gamma_j\gamma_i = 2\delta_{ij} \tag{4.1.1}$$

Majorana fermions always come in pairs in a finite physical system, and $2N$ of them with (4.1.1) generate a real Clifford algebra [Lawson and Michelsohn, 1989]. $\gamma_1, \ldots, \gamma_{2N}$ together with the fermion parity operator

$$(-1)^F = (-i)^N \gamma_1 \ldots \gamma_{2N} \tag{4.1.2}$$

can be represented by a maximal set of anticommuting operators in $su(2)^{\otimes N}$, and therefore act on a 2^N dimensional Fock space. In a closed system, the fermion parity cannot change, and the Fock space decomposes into even and odd irreducible components, each being 2^{N-1} dimenional.

The Fock states can also be understood as occupation states $\prod_{j=1}^{N}(c_j^\dagger)^{n_j}|0\rangle$, for $n_j = 0$ or 1, of Dirac fermions

$$c_j = \frac{1}{2}(\gamma_{2j-1} + i\gamma_{2j}) \tag{4.1.3}$$

with fermion parity $(-1)^F = (-1)^{\sum_{j=1}^{N} n_j}$. A two-level quantum state $a_0|0\rangle + a_1|1\rangle$ is therefore non-locally carried by a pair of Majorana fermions that could be spatially far apart and cannot be measured independently [Kitaev, 2003]. The Majorana

excitations considered in this section are distinct point-like low energy quasiparticles inside a bulk gapped material. The wavefunction overlapping and electron tunneling between any two quasiparticles is exponentially suppressed in separation distance by the finite bulk excitation energy gap. The Fock space dimension 2^{N-1} thus corresponds to the ground state degeneracy of such a closed system in the infinite separation/gap limit.

4.1.1 Majorana fermions in low dimensions

The appearance of low energy Majorana bound state can be most easily illustrated by a chiral 1D superconductor (a.k.a. Kitaev's Majorana chain) [Kitaev, 2001]. There is a positive excitation energy gap unless the Fermi energy μ and hopping t

$$H_{(A)} - \mu N = \sum_j t(c_j^\dagger c_{j+1} + h.c.) - \frac{\mu}{2}(c_j^\dagger c_j - c_j c_j^\dagger) - \Delta(c_j c_{j+1} + h.c.) \qquad (4.1.4)$$

$$H_{(B)} - \mu N = \sum_j iJ\gamma_{2j}\gamma_{2j+1} + ih\gamma_{2j-1}\gamma_{2j} \qquad (4.1.5)$$

Figure 4.1: (A) 1D chiral p-wave superconductor with hopping t, pairing Δ and Fermi energy μ. (B) Kitaev's Majorana chain.

satisfy $|\mu| = 2|t|$. The relative magnitude of hopping and Fermi energy thus defines two superconducting phases. The parameters can be tuned continuously in the same phase without closing the excitation gap. In the limit $|\Delta| \to |t|$, the model changes to the Kitaev's Majorana chain using (4.1.3), for $|h| = |\mu|/2$ and $|J| = |t|$.[2]

---

[2]The Kitaev's Majorana chain (4.1.5) is identical to the quantum Ising model in 1D (or equivaletly the classical Ising model in 2D by transfer matrix [Baxter, 1982])

$$H = -J\sum_j \sigma_x^j \sigma_x^{j+1} - h\sum_j \sigma_z^j \qquad (4.1.6)$$

The Kitaev's chain (or the 1D superconductor) is in the topological phase when $|h| < |J|$ (or $|\mu| < 2|t|$). Just like the Su-Schrieffer-Heeger model [1980] (fig.1.7), there are dangling states left at the terminals of a finite chain (see fig.4.2). The
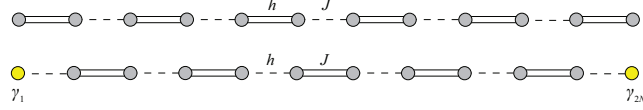
Figure 4.2: (Top) Trivial phase for $|h| > |J|$. (Bottom) Topological phase for $|h| < |J|$ with Majorana bound states (yellow) at the ends.

Majorana bound states are zero energy modes of the Hamiltonian in the thermodynamic limit. They are protected from a particle-hole (or charge conjugation) symmetry and they represent the ground states of the system. A Dirac fermion can be constructed by the pair of Majorana end states $c = (\gamma_1 + i\gamma_{2N})/2$ and defines a two level system. The two-fold degeneracy corresponds to the two ordered ground states of the Ising model.
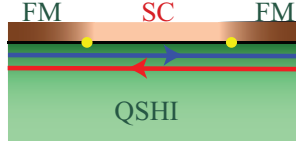
Figure 4.3: Majorana fermions (yellow) at quantum spin Hall insulator (QSHI) - ferromagnet (FM) - superconductor (SC) interface [Fu and Kane, 2009a,b; Teo and Kane, 2010a,b].

Majorana fermions have Ising non-Abelian statistics in $(2+1)$-dimensions [Nayak and Wilczek, 1996; Ivanov, 2001]. Unlike ordinary bosons, fermions or Abelian anyons [Wilczek, 1982; Wilczek and Zee, 1983; Halperin, 1984; Arovas, Schrieffer and Wilczek, 1984], the exchange operator between a pair of non-Abelian anyons is represented by a unitary matrix that does not commute with exchange operators of other pairs of anyons. This means that different orderings of exchange operations in general give different final states.

---

by the Jordan-Wigner transformation $\gamma_{2j-1} = \sigma_x^j \prod_{i<j} \sigma_z^i$ and $\gamma_{2j} = \sigma_y^j \prod_{i<j} \sigma_z^i$. This explains the alternative terminology Ising anyons for Majorana fermions.
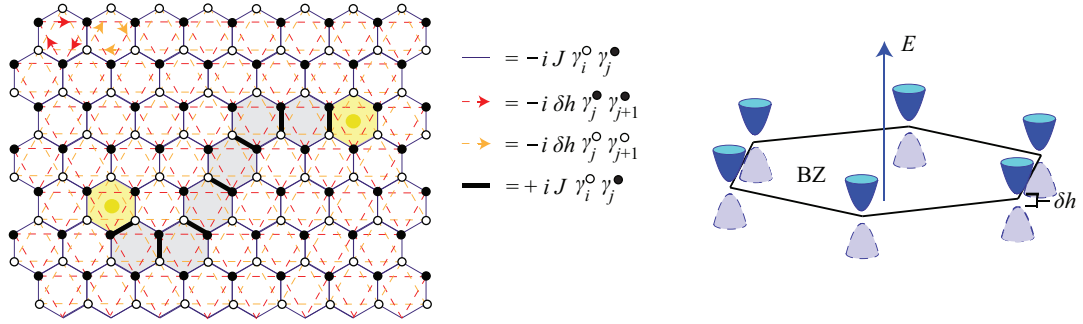
Figure 4.4: Fermionization of the Kitaev's honeycomb B-model with a magnetic field [Kitaev, 2006]. $\gamma_i^\circ, \gamma_j^\bullet$ are Majorana operators on the A,B-sublattice. Majorana bound state (orange dots) are located at plaquettes with π-flux (yellow hexagons). Solid black lines represent (-1)'s of the $\mathbb{Z}_2$-gauge field, and they form a string connecting the Majorana fermions (shaded hexagons).
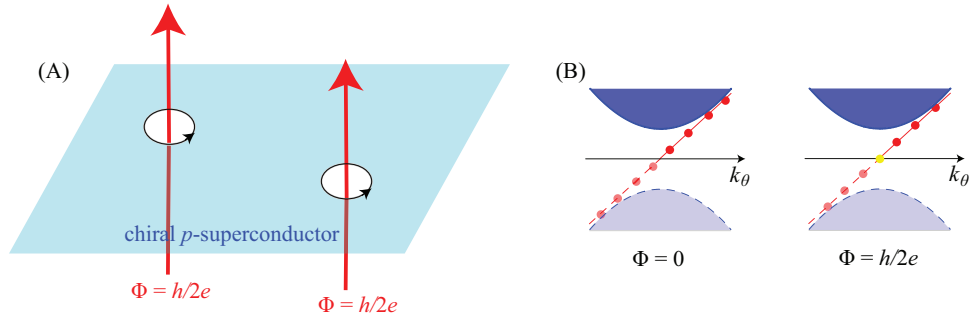
Figure 4.5: (A) Chiral p-wave superconductor with $h/2e$ vortices [Volovik, 1999; Read and Green, 2000]. (B) Discrete energy spectrum of chiral Majorana mode along a circular hole boundary. Flux $\Phi = h/2e$ shifts the states by π/N and a Majorana bound state (yellow) sits exactly at zero energy.

Ising anyons are non-Abelian anyons so that when one circles around the other, both of them accumulate a minus sign. This means that the exchange operators is not simply interchanging 2 Majorana operators but it also changes the sign of one of them.

The minus sign of the exchange operator (4.1.8) comes from the π-flux associated with each Majorana fermions. A Berry phase of $e^{i\pi} = -1$ is accumulated when one adiabatically circles around the other. For example, the Majorana fermions in the Kitaev's honeycomb B-model [Kitaev, 2006] are connected by strings of (-1)-links (fig.4.4), and Majorana fermions appear at $h/2e$ vortex cores of a superconductor
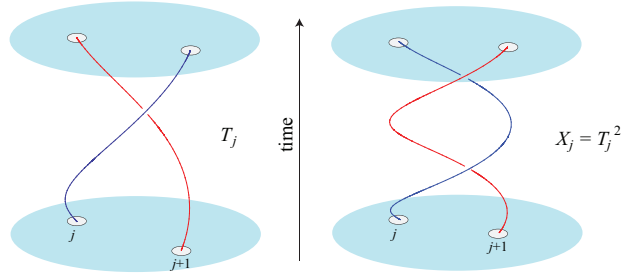
$$T_j = \exp\left(\pi\gamma_j\gamma_{j+1}/4\right), \quad X_j = T_j^2 = \gamma_j\gamma_{j+1}, \quad \text{(up to a } U(1) \text{ phase)} \tag{4.1.7}$$

$$\begin{cases} T_j\gamma_j T_j^{-1} = -\gamma_{j+1} \\ T_j\gamma_{j+1}T_j^{-1} = \gamma_j \end{cases} \quad \text{and} \quad \begin{cases} X_j\gamma_j X_j^{-1} = -\gamma_j \\ X_j\gamma_{j+1}X_j^{-1} = -\gamma_{j+1} \end{cases} \tag{4.1.8}$$

Figure 4.6: Exchange and double exchange operators of Ising anyons.

[Volovik, 1999; Read and Green, 2000] with odd Chern invariant (fig.4.5). The π-Berry phase is topologically protected by non-trivial *braiding* of the quasiparticles.
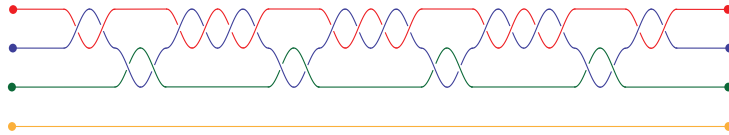
Figure 4.7: Example of a braid.

The exchange operators $T_j = \exp(\pi\gamma_j\gamma_{j+1}/4)$, for $j = 1, \ldots, 2N-1$, generates a collection of non-commuting unitary operations. They form a representation of the *Braid group*[3] $\mathcal{B}_{2N}$ with the relation

$$T_jT_{j+1}T_j = T_{j+1}T_jT_{j+1} \quad \text{and} \quad T_iT_j = T_jT_i, \quad \text{for } |i - j| \geq 2 \tag{4.1.9}$$

The topology of braiding is unique in $(2+1)$ dimensions. All braids with the same end points are topologically equivalent in higher dimensions as there are more space

---

[3]Defined as the fundamental group $\mathcal{B}_{2N} = \pi_1(\mathcal{C}_{2N}^{2D})$ of the configuration space $\mathcal{C}_{2N}^{2D}$ of the positions of $2N$ distinct indistinguishable points in $\mathbb{R}^2$. It is the group generated by T_j, $j = 1, \ldots, 2N-1$, with the relation (4.1.9), and has infinitely many elements.

for them to continuously deform into one another without crossing. So the braid group in higher dimensions are just the symmetric group S_{2N} of permutation. However, the principle of locality suggests that the Hilbert space must decompose into tensor products of local Hilbert spaces associated to individual particles [Doplicher, Haag and Roberts, 1971, 1973], essentially bosons and fermions. This forbids for instance the non-local storage of quantum information between two spatially separated Majorana fermions. This seemingly prohibits Ising non-Abelian statistics in higher dimensions, but can in fact be circumvented by a more careful look at the configuration space of Majorana excitations [Teo and Kane, 2010a; Freedman *et.al.*, 2011].

4.2   Topological point defect in 3D

Before the discovery of topological insulators, non-Abelian anyons in electronic systems are mostly investigated in 2D chiral p-wave superconductor [Volovik, 1999; Read and Green, 2000; Ivanov, 2001] or various fractional quantum Hall states, such as the Pfaffian state [Moore and Read, 1991], Read-Rezayi states [Read and Rezayi, 1999] and the non-Abelian spin-singlet state [Ardonne and Schoutens, 1999]. While strontium ruthenate Sr_2RuO_4 is a chiral p-wave superconductor [Das Sarma, Nayak and Tewari, 2006], experimentally creating Majorana bound excitations on the material involves complication of stablizing half vortices on a thin film [Jang *et.al.*, 2011]. The non-Abelian fraction quantum Hall states on the other hand requires a very clean sample with high mobility and exceptionally low temperature. Topological insulator (TI) - superconductor (SC) heterostructures could in theory provide

an alternative arena for Majorana excitations and manipulations that circumvent these difficulties.

In a BCS mean field description of superconductivity, excitations are understood by the Bogoliubov de Gennes (BdG) Hamiltonian [De Gennes, 1966]

$$H - \mu N = \sum_{k>0} \vec{\xi}^{\dagger}_{-\mathbf{k}} H_{BdG}(\mathbf{k}) \vec{\xi}_{\mathbf{k}} \qquad (4.2.1)$$

under the Nambu basis $\vec{\xi}_{\mathbf{k}} = (c_{\mathbf{k}}, c^{\dagger}_{-\mathbf{k}})^T$ that incorporates both creation and annihilation operators. An artificial particle-hole (PH) symmetry

$$\Xi^{-1} H_{BdG}(\mathbf{k}) \Xi = -H_{BdG}(-\mathbf{k}) \qquad (4.2.2)$$

always arises in the BdG Hamiltonian due to a double counting of degrees of freedom: $\Xi : c_{\mathbf{k}} \leftrightarrow c^{\dagger}_{-\mathbf{k}}$. As a charge conjugation symmetry, the PH operator Ξ must be antiunitary, and due to the fermionic nature of electrons, $\Xi^2 = +1$.

The parameters in the BdG Hamiltonian may undergo a slow spatial modulation, say the paring term $\Delta e^{i\phi} c_j c_j + h.c.$ away from a vortex line or in the presence of a spatially varying Fermi energy $\mu(\mathbf{r})$. The spatial modulation is treated as adiabatic perturbation so that the superconductor still maintains a *local* translation symmetry, and crystal momentum $\mathbf{k}$ is defined semi-classically. Since PH symmetry is local, it requires

$$\Xi^{-1} H_{BdG}(\mathbf{k}, \mathbf{r}) \Xi = -H_{BdG}(-\mathbf{k}, \mathbf{r}) \qquad (4.2.3)$$

A notion of *topological defect*[4] can be defined by how the semi-classical BdG

---

[4]For more, see Teo and Kane [2010b] or section 5.3.2 in the next chapter.

Hamiltonian behaves as a function of distance $\mathbf{r}$ wrapping around the defect. Given any closed surface Σ in real 3D space, the number of Majorana excitations enclosed (mod 2) is given by a $\mathbb{Z}_2$ Chern-Simons invariant [Teo and Kane, 2010a,b].

$$\nu \equiv \frac{2}{3!} \left(\frac{i}{2\pi} \right)^3 \int_{BZ \times \Sigma} \mathcal{Q}_5 \quad (\text{mod } 2) \tag{4.2.4}$$

where the Chern-Simons 5-form (omitting wedge products $\wedge$ between $\mathcal{A}$'s) [Chern and Simons, 1974; Nakahara, 1990]

$$\mathcal{Q}_5 = \text{Tr} \left(\mathcal{A} \wedge (d\mathcal{A})^2 + \frac{3}{2}\mathcal{A}^3 \wedge d\mathcal{A} + \frac{3}{5}\mathcal{A}^5 \right) \tag{4.2.5}$$

is defined by the Berry connection

$$\mathcal{A}_{mn} = \langle u_m(\mathbf{k}, \mathbf{r})|du_n(\mathbf{k}, \mathbf{r})\rangle = \langle u_m(\mathbf{k}, \mathbf{r})|\partial_\mathbf{r} u_n(\mathbf{k}, \mathbf{r})\rangle \cdot d\mathbf{r} + \langle u_m(\mathbf{k}, \mathbf{r})|\partial_\mathbf{k} u_n(\mathbf{k}, \mathbf{r})\rangle \cdot d\mathbf{k}$$

$$\tag{4.2.6}$$

constructed from the "valence" band $u_m(\mathbf{k}, \mathbf{r})$ of the BdG Hamiltonian $H_{BdG}(\mathbf{k}, \mathbf{r})$. PH symmetry guarantees that the Chern-Simons integral (4.2.4) is an integer. And because of the factor of 2 in front of the integration, a large gauge transformation can only alter the invariant by an integral multiple of 2.

The correlation of the $\mathbb{Z}_2$ invariant and the existence of a Majorana bound state is a generalization of the bulk boundary correspondence. It can be proven rigorously using heat kernel techniques in showing the Atiyah-Singer index theorem [Atiyah and Singer, 1968; Nakahara, 1990; Berline, Getzler and Vergne, 2004]. A verification of the correspondence is given in the following using a minimal 8 band model that capture the low energy physics at a TI-SC interface. The interface should not
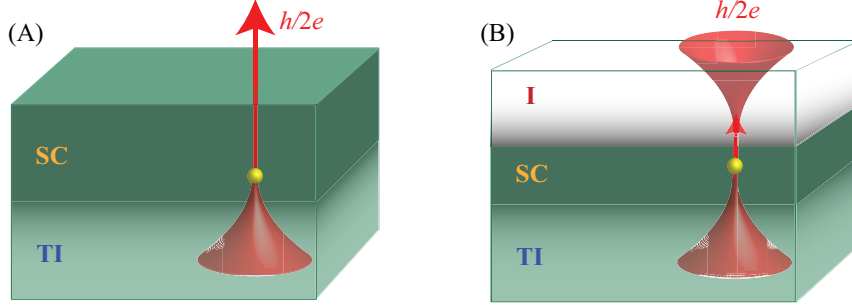
Figure 4.8: Majorana fermions (yellow) in 3D. (A) A $h/2e$ flux vortex across a topological insulator (TI) - type-II superconductor (SC) interface. (B) A flim of superconductor (SC) sandwhiched between a topological insulator (TI) and a trivial insulator (I). Majorana zero mode is 3-dimensionally located somewhere in the superconducting film.

be treated as a distinct surface. Cooper pairs can tunnel a certain depth into the insulator, and the BdG Hamiltonian remains gapped throughout. If a $h/2e$ flux passes across the interface, low energy excitations near the vortex core can be qualitatively studied by the continuum BdG model

$$H(\mathbf{k}, \mathbf{r}) = \left(\hbar v_f \mathbf{k} \cdot \vec{\sigma} \mu_x + (m(\mathbf{r}) + \epsilon k^2) \mu_z\right)\tau_z + \Delta_x(\mathbf{r})\tau_x + \Delta_y(\mathbf{r})\tau_y \qquad (4.2.7)$$

Here the τ matrices act on the Nambu basis, σ's and μ's are spin-orbital coupled degrees of freedom. $\Delta(\mathbf{r}) = \Delta_x(\mathbf{r}) + i\Delta_x(\mathbf{r})$ is the SC pairing parameter and it winds 2π around the vortex in the superconductor. The coefficient of τ_z in the first term is a 4-band model describing a topological insulator (TI) to trivial insulator (I) band inversion, where the mass term $m(\mathbf{r})$ changes sign along the vertical axis.[5] The model has PH symmetry according to $\Xi = \tau_y \sigma_y K$ and the vortex pairing term $\Delta_y(\mathbf{r})\tau_y$ breaks time reversal symmetry $\Theta = i\sigma_y K$.[6]

If we name the Dirac matrices by $\vec{\gamma} = \vec{\sigma}\mu_x\tau_z$ and $(\Gamma_x, \Gamma_y, \Gamma_z) = (\tau_x, \tau_y, \mu_z\tau_z)$, the

---

[5] The second order term ϵk^2 compactifies the Brillouin zone into a sphere $\mathbb{S}^3 = \mathbb{R}^3 \cup \{\infty\}$ and the relative signs between m and ϵ determines the topological phase of the insulator. In the context of defect, it is not essential to know whether the top or bottom insulator (fig.4.8(B)) is topological as long as there is a phase transition across. The k^2 term therefore can be dropped.

[6] There is an artificial chiral symmetry $\Pi = \mu_y\tau_z$ in the model but this would be killed by a non-zero Fermi energy term $\mu\tau_z$.
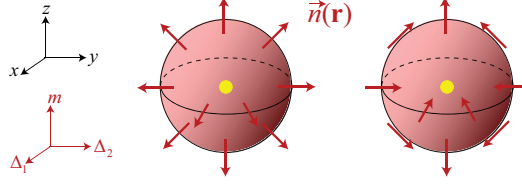
Figure 4.9: Unit winding degree hedgehogs in real space and Majorana bound states (yellow) at point defects.

BdG Hamiltonian has the simple form

$$H(\mathbf{k}, \mathbf{r}) = \hbar v_f \mathbf{k} \cdot \vec{\gamma} + \vec{n}(\mathbf{r}) \cdot \vec{\Gamma}, \quad \vec{n}(\mathbf{r}) = (\Delta_x(\mathbf{r}), \Delta_y(\mathbf{r}), m(\mathbf{r})) \tag{4.2.8}$$

Here the 3-component vector field $\vec{n}(\mathbf{r})$ controls the spatial modulation of the Hamiltonian. Its magnitude determines the energy gap for *local* bulk excitation. The singular points (such as sources and drains) of the vector field are the point defects of the system, and acquire *hedgehog* configurations (fig.4.9). The topology of a hedgehog is given by the Chern-Simons integral (4.2.4) over an enclosing sphere $\Sigma = \mathbb{S}^2$, which simplifies to the winding integer

$$\nu = \frac{1}{8\pi} \int_{\mathbb{S}^2} \boldsymbol{\omega} \cdot d\mathbf{A}, \quad \omega_i = \frac{1}{8\pi} \hat{n} \cdot \partial_j \hat{n} \times \partial_k \hat{n} \tag{4.2.9}$$

where $\hat{n} = \vec{n}/|n|$ is the unit direction. Treating $\hat{n}(\mathbf{r})$ as a map sending a point in the enclosing surface sphere $\Sigma = \mathbb{S}^2$ in real space to a unit vector in the parameter space, the winding degree (4.2.9) counts the number of times $\hat{n} : \mathbb{S}^2 \to \mathbb{S}^2$ wraps the first sphere around the second one. Mathematically, it integrally represents the hedgehog $\hat{n}$ in the *homotopy group* [Hatcher, 2002] $\pi_2(\mathbb{S}^2) \cong \mathbb{Z}$ of topologically equivalent maps $\mathbb{S}^2 \to \mathbb{S}^2$.[7]

---

[7]A general isotropic Dirac-type BdG Hamiltonian has the following form

$$H(\mathbf{k}, \mathbf{r}) = \hbar v_f \mathbf{k} \cdot \vec{\gamma} + m\Gamma(\mathbf{r}) \tag{4.2.10}$$

where the mass term $\Gamma(\mathbf{r})$ is a hermitian matrix anticommutes with the Dirac matrices $\gamma_x, \gamma_y, \gamma_z$ and the PH operator Ξ. All matrices are $2n \times 2n$ for a $2n$-band model. Symmetry requires $[\Xi, \gamma_i] = 0$, and the exitation gap is

In particular, a unit winding linear hedgehog would take the form of a monopole $n_i(\mathbf{r}) = M_{ij}r_j$, where the non-singular matrix M decomposes into

$$M = O_1^T \begin{pmatrix} M_1 & 0 & 0 \\ 0 & M_2 & 0 \\ 0 & 0 & M_3 \end{pmatrix} O_2 \tag{4.2.13}$$

for some positive numbers M_a and orthogonal matrices O_1, O_2 that diagonalize MM^T, M^TM. The semi-classical description breaks down at the center of the monopole. By replacing $\mathbf{k} \leftrightarrow -i\nabla$, the BdG Hamiltonian (4.2.8) becomes a Dirac operator

$$\hat{H} = -i\hbar v_f \vec{\gamma} \cdot \nabla + \Gamma_i M_{ij}r_j \tag{4.2.14}$$

Eigenstates can be solve exactly by noticing that the square of the operator is a 3D harmonic oscillator (a.k.a. Laplacian)

$$\begin{aligned} \hat{H}^2 &= -\hbar^2 v_f^2 \nabla^2 + |M\mathbf{r}|^2 + i\hbar v_f \Gamma_i M_{ij}\gamma_j & (4.2.15) \\ &= \hbar v_f \sum_{a=1,2,3} M_a(2n_a + 1 - \xi_a) & (4.2.16) \end{aligned}$$

where $n_a = 0, 1, 2, \dots$ are the oscillator quantum numbers and $\xi_a = \pm 1$ are the eigenvalues of the commuting Dirac matrices $i(O_{2ai}\gamma_i)(O_{1aj}\Gamma_j)$. A unique zero energy mode is given by $n_a = 0$ and $\xi_a = 1$ and describes a low energy Majorana

finite if $\Gamma(\mathbf{r})$ is non-singular. Choose a gauge so that $\Xi = \sigma_y \tau_y K$, $\vec{\gamma} = \vec{\sigma}\tau_z$. The mass term satisfies

$$\Gamma(\mathbf{r}) = \mathbb{1} \otimes (\tau_x \otimes g_x(\mathbf{r}) + \tau_y \otimes g_y(\mathbf{r})), \quad g = g^T = g_x + ig_y \in GL(n/2; \mathbb{C}) \tag{4.2.11}$$

A symmetric invertible complex matrix can be decomposed into $g = G^T G = G^T O^T OG$, where $G \in GL(n/2; \mathbb{C})$ is defined up to $O \in O(n/2; \mathbb{C})$. Hence the parameter space for the mass term is the symmetric space [Cartan, 1926]

$$\Gamma(\mathbf{r}) \in \frac{GL(n/2; \mathbb{C})}{O(n; \mathbb{C})} \simeq \frac{U(n/2)}{O(n/2)} \tag{4.2.12}$$

which has stable homotopy groups [Bott, 1956] $\pi_1(U/O) = \mathbb{Z}$, $\pi_2(U/O) = \mathbb{Z}_2$ corresponding to topological line and point defects.

bound state.

4.3   Non-abelian statistics in 3D

Low energy Majorana bound states in 3D can be understood by the Dirac-type model Hamiltonian (4.2.8) (or the more general (4.2.10)) as point defects associated with a hedgehogs of the vector field $\vec{n}(\mathbf{r})$ (resp. the mass parameter $m\Gamma(\mathbf{r})$). The configuration space of the system thus not only consists of the locations of the $2N$ distinct indistinguishable Majorana fermions but also the non-singular vector field $\vec{n}(\mathbf{r})$ throughout real space. It is a feature that leads to non-Abelian statistics [Teo and Kane, 2010a; Freedman *et.al.*, 2011].

4.3.1   Fermion parity pump

The origin of Ising statistics is a minus sign in the exchange operator (4.1.8). The minus sign can be identified with a π-Berry phase of a 360° rotation. Consider an adiabatic evolution

$$H_\phi(\mathbf{k}, \mathbf{r}) = H_0(\mathbf{k}, R_\phi\mathbf{r}) = U_\phi H_0(\mathbf{k}, \mathbf{r})U_\phi^\dagger \tag{4.3.1}$$

where U_ϕ is a unitary (rotation) operator on the Hilbert space representing R_ϕ, and $\phi = \phi(t)$ is some adiabatic (angle) evolution. PH symmetry requires $[\Xi, U_\phi] = 0$ since rotation never mixes $c, c^\dagger$. Suppose $|\psi_0\rangle$ is a non-degenerate Majorana zero mode with its phase fixed so that $\Xi|\psi_0\rangle = +|\psi_0\rangle$.[8] PH symmetry then ensures

---

[8]This is possible only when $\Xi^2 = +1$ so that there is no Kramer's degeneracy in the zero energy eigenspace. The PH operator Ξ then form a *real* structure on zero modes.
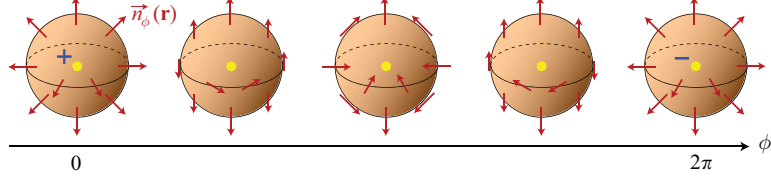
Figure 4.10: Skyrmion-like evolution of the vector field $\vec{n}_\phi(\mathbf{r}) = MR_\phi \mathbf{r}$. The Majorana zero mode (yellow) acquires a minus sign after a cycle.

$\Xi|\psi_\phi\rangle = \Xi U_\phi|\psi_0\rangle = +|\psi_\phi\rangle$ throughout the cycle. And therefore the zero mode can only accumulate a discrete $\mathbb{Z}_2$ phase after one complete cycle,

$$|\psi_{2\pi}\rangle = \pm|\psi_0\rangle \qquad (4.3.2)$$

if it is the only Majorana zero mode in the system. The π-Berry phase is topological in the sense that it cannot be accumulated from a trivial contractible adiabatic cycle. The sign in eq.(4.3.2) then signifies a topological $\mathbb{Z}_2$ classification for PH-symmetric adiabatic cycles, also known as *fermion parity pumps*.

To show that a π-Berry phase can be achieved, we consider a 360° rotation of the hedgehog $\vec{n}_\phi(\mathbf{r}) = R_\phi \vec{n}_0(\mathbf{r}) = MR_\phi \mathbf{r}$. Say R_ϕ rotates about the z-axis. The minimal 8-band model (4.2.8) evolves according to

$$H_\phi = -i\hbar v_f \vec{\gamma} \cdot \nabla + M\vec{\Gamma} \cdot R_\phi \mathbf{r} = U_\phi H_0 U_\phi^\dagger \qquad (4.3.3)$$

for $U_\phi = \exp(\phi\Gamma_x\Gamma_y/2)$. Thus $|\psi_{2\pi}\rangle = U_{2\pi}|\psi_0\rangle = -|\psi_0\rangle$.

The π-Berry phase is topological protected by the skyrmion-like evolution the vector field $\vec{n}_\phi(\mathbf{r})$ undergoes (fig.4.10). Treat the vector field as a map $\hat{n} = \vec{n}/|n|$: $\mathbb{S}^1_\phi \times \mathbb{S}^2_\mathbf{r} \to \mathbb{S}^2$ where the angle parameter ϕ lives in $\mathbb{S}^1_\phi$, $\mathbb{S}^2_\mathbf{r}$ is a 2-sphere in real space enclosing the Majorana bound state, the target $\mathbb{S}^2$ parametrizes the uni direction

51

of $\vec{n} = (\Delta_1, \Delta_2, m)$. We therefore seek the topological classification of *homotopy classes* of general maps from $\mathbb{S}^1 \times \mathbb{S}^2$ to $\mathbb{S}^2$. This has been studied mathematically by Pontrjagin [1941].

The existence of Majorana bound state requies the winding degree of the hedgehog configuration $\hat{n}_\phi : \mathbb{S}^2_{\mathbf{r}} \to \mathbb{S}^2$ to be odd, say $\deg(\hat{n}_\phi) = p$, for all ϕ. Pontryagin asserted that the classification with this degree constraint is

$$[\mathbb{S}^1 \times \mathbb{S}^2, \mathbb{S}^2]_{\deg=p} = \mathbb{Z}_{2p} \tag{4.3.4}$$

For unit degree, we may well assume the vector field takes the linear form $\hat{n}_\phi(\mathbf{r}) = R_\phi \hat{r}$, where R_ϕ is some rotation in $SO(3)$.[9] Then the topological classification is simply

$$[\mathbb{S}^1 \times \mathbb{S}^2, \mathbb{S}^2]_{\deg=1} = \pi_1(SO(3)) = \mathbb{Z}_2 \tag{4.3.7}$$

which also accounts for the minus sign in $360°$ rotation of a fermion. We therefore see that the skyrmion cycle is a topologically non-trivial evolution.[10]

There must be an even number of Majorana fermions in a closed system and any rotation of the vector field must change the signs of a pair or pairs of them. An example is shown in fig.4.11 for a single pair. In order for the fermion parity pump

---

[9]It is easiest to understand in the degree one scenario by the following conceptual but non-rigorous series of identifications.
$$\mathrm{Map}(\mathbb{S}^1 \times \mathbb{S}^2, \mathbb{S}^2)_{\deg=1} \sim \mathrm{Map}\left(\mathbb{S}^1, \mathrm{Map}(\mathbb{S}^2, \mathbb{S}^2)_{\deg=1}\right) \sim \mathrm{Map}(\mathbb{S}^1, SO(3)) \tag{4.3.5}$$
Here the first equality is simply saying $\hat{n}_\phi$ is a ϕ-parametrized collection of maps from $\mathbb{S}^2$ to $\mathbb{S}^2$. The second equality can be understood by assuming an analytic structure to the maps.
$$\mathrm{Map}(\mathbb{S}^2, \mathbb{S}^2)_{\deg=1} \sim \mathrm{Aut}(\mathbb{S}^2) = SL(2; \mathbb{C}) \simeq SO(3) \tag{4.3.6}$$
where $\mathrm{Aut}(\mathbb{S}^2)$ are complex analytic automorphisms on the Riemann sphere, which deformation retract onto the rotation group $SO(3)$.

[10]This topology is not specific to an 8-band model. In a general Dirac-type model, the mass term $m\Gamma_\phi(\mathbf{r})$ in eq.(4.2.10) can undergoes a non-trivial cycle stably classified by [Bott, 1956; Freedman *et.al.*, 2011]
$$[\mathbb{S}^1 \times \mathbb{S}^2, U/O] \sim \pi_3(U/O) = \mathbb{Z}_2 \tag{4.3.8}$$
In fact, the adiabatic cycle falls in the $\mathbb{Z}_2$-classification of topological $\delta = 0$ defects of class D [Teo and Kane, 2010b].
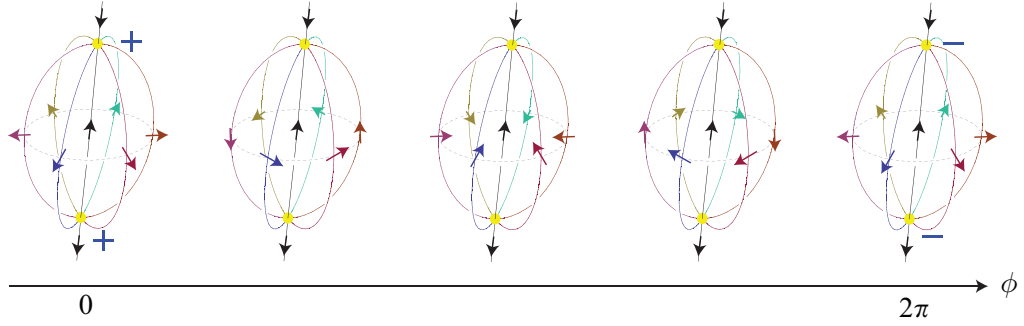
$$\vec{n}_\phi(\mathbf{r}) \propto \left(x\cos\phi + y\sin\phi, -x\sin\phi + y\cos\phi, \frac{1-r^2}{2} \right), \quad \vec{n}_\phi(r \to \infty) \propto \mathbf{e}_3$$

Figure 4.11: Skyrmion evolution of a pair of hedgehogs in a closed system. $\vec{n}_\phi$ points at a constant direction (colored arrow) along a colored line. Both Majorana bound states (yellow) at $\mathbf{r} = (0, 0, \pm 1)$ pick up minus signs.

to be physically measurable, there must be at least two pairs of Majorana fermions. So that two *local* fermion parity operators can be defined, and a single (or an odd number of) fermion is pumped across the two spatially separated regions. Say we have $2N_L$ Majorana bound states on one side and $2N_R$ on the other. Local fermion parity operators are defined according to

$$(-1)^{F_a} = (-i)^{N_a} \gamma_1^a \dots \gamma_{2N_a}^a, \quad a = L/R \tag{4.3.9}$$

The two sides are interconnected throught the vector field $\vec{n}(\mathbf{r})$, and suppose a certain evolution changes the signs of a pair of Majorana fermions γ_i^L, γ_j^R between the two sides. The unitary operator is represented by

$$X = \gamma_i^L \gamma_j^R \quad \text{(up to a } U(1) \text{ phase)} \tag{4.3.10}$$

so that $X\gamma_{i'}^L X^\dagger = (-1)^{\delta_{ii'}} \gamma_{i'}^L$ and $X\gamma_{j'}^R X^\dagger = (-1)^{\delta_{jj'}} \gamma_{j'}^R$. X *anticommutes* with the local fermion parity operators but *commutes* with the total fermion parity $(-1)^{F_L+F_R}$. This shows that after a cycle, a single (or an odd number of) fermion
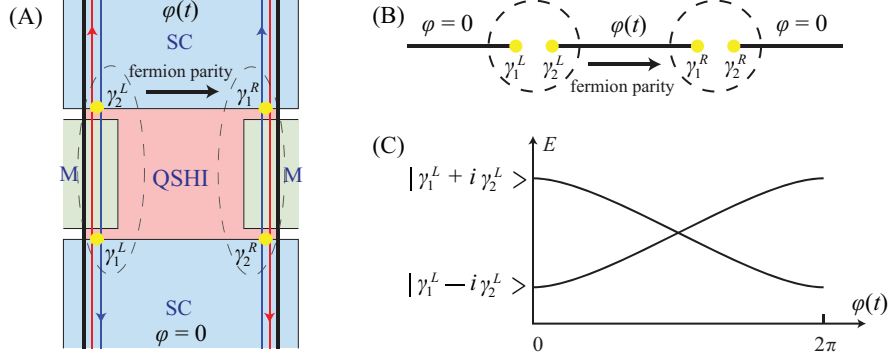
Figure 4.12: Fermion parity pumps controlled by phase $\varphi(t)$ of superconductors. (A) Quantum spin Hall insulator (QSHI) - magnetic insulator (M) - superconductor (SC) interface. Left and right edges are similar to fig.4.3, each bounds two Majorana bound states (orange). γ_2^L and γ_1^R changes signs after a cycle of relative SC phase evolution. A single (or an odd number of) fermion is pump between the left and right systems. (B) Phase evolution of SC chain. $H_{BdG}(k, \varphi) = (t \cos k - \mu)\tau_z + \Delta \sin k(\cos \varphi \tau_x + \sin \varphi \tau_y)$. (C) Energy levels associated with electron tunneling across the non-superconducting gaps between γ_1^a and γ_2^a, for $a = L, R$. Crossing is protected by topology of the adiabatic cycle (see eq.(5.3.25) in section 5.3.3).

has traveled across the two sides through the bulk.

4.3.2 Braidless operations (a.k.a. double exchanges)

Braidless operations are generalizations of fermion parity pumps [Teo and Kane, 2010a; Bonderson, Freedman and Nayak, 2008]. Given a system with $2N$ distinct Majorana bound states, each associated with a hedgehog in the vector field $\vec{n}(\mathbf{r})$ (or point defect in the mass term $m\Gamma(\mathbf{r})$), braidless operations are topological non-trivial adiabatic cycles $\vec{n}_t(\mathbf{r})$ *without moving* the hedgehogs, i.e. stationary Majorana fermions. Fermions are pumped in between pairs of Majorana bound states and altering parity of local pairs.

In a closed system with fixed boundary condition $\vec{n}(r \to \infty) = \text{constant}$, the topology of the vector field $\vec{n}(\mathbf{r})$ is specified on a bouquet of spheres $\wedge^{2N-1}\mathbb{S}^2$ (see fig.4.13), each enclosing one Majorana fermion. The topology around the remaining
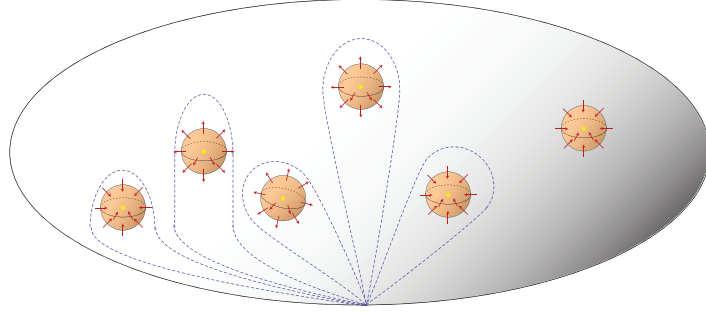
Figure 4.13: $2N$ Majorana fermions in a closed system with fixed boundary condition. Blue dotted loops denote a bouquet of $(2N-1)$ number of spheres $\wedge^{2N-1}\mathbb{S}^2$ enclosing all but one hedgehogs.

unenclosed one is specified by the rest and the boundary condition. The configuration space is therefore[11]

$$\text{Map}\left(\wedge^{2N-1}\mathbb{S}^2,\mathbb{S}^2\right) = \text{Map}\left(\mathbb{S}^2,\mathbb{S}^2\right)_{\text{deg}=1}^{2N-1} \sim SO(3)^{2N-1} \tag{4.3.12}$$

Adiabatic cycles $\vec{n}_t(\mathbf{r})$ (or mass term $m\Gamma_t(\mathbf{r})$) are topologically classified by the fundamental group

$$\pi_1(SO^{2N-1}) = \pi_1(SO)^{2N-1} = (\mathbb{Z}_2)^{2N-1} \tag{4.3.13}$$

It is more convenient to represent the group as the even part of $(\mathbb{Z}_2)^{2N}$, which is generated by $x_{ij} = (x_{ij}^1, \ldots, x_{ij}^{2N})$, $x_{ij}^l = (-1)^{\delta_{il}+\delta_{jl}}$. They have the Abelian group relation

$$x_{ij}x_{kl} = x_{kl}x_{ij}, \quad x_{ij}^2 = 1 \tag{4.3.14}$$

Physically, x_{ij} represents the adiabatic cycle where the ith and j^{th} Majorana fermions undergo $360°$ rotation similar to fig.4.11. Hence, both of them change

---

[11]A general Dirac-type model would have configuration space

$$\text{Map}\left(\wedge^{2N-1}\mathbb{S}^2, U/O\right) = \text{Map}\left(\mathbb{S}^2, U/O\right)_{\text{deg}=1} \sim SO^{2N-1} \tag{4.3.11}$$
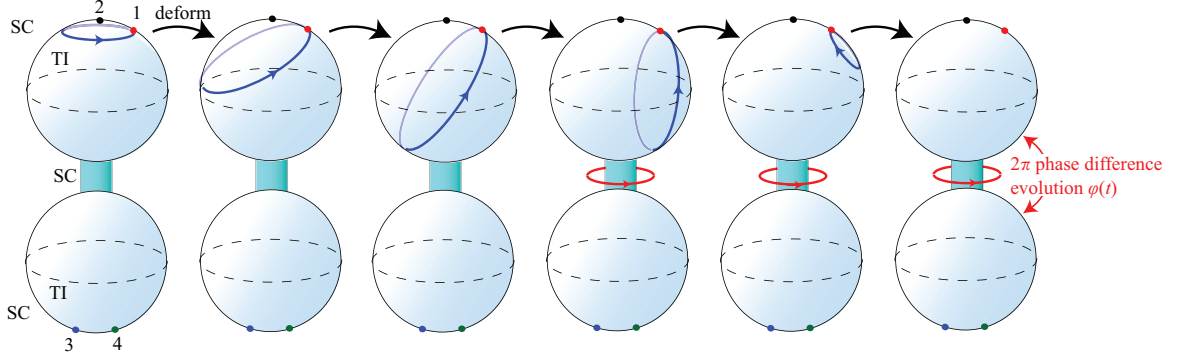
Figure 4.14: Deforming a double exchange T_{12}^2 (far left) into a braidless operation X_{12} (far right) in 3D. A single (or odd number of) fermion parity is pumped between the top and bottom superconducting spheres.

signs and it is represented by the unitary operator

$$X_{ij} = \gamma_i \gamma_j \quad \text{(up to a } U(1) \text{ phase)} \tag{4.3.15}$$

so that $X_{ij} \gamma_l X_{ij}^\dagger = x_{ij}^l \gamma_l = (-1)^{\delta_{il} + \delta_{jl}} \gamma_l$. Interestingly, they form a non-commuting set of operators which differ from (4.3.14) by a sign

$$X_{ij} X_{kl} = (-1)^{\delta_{ik} + \delta_{jk} + \delta_{il} + \delta_{jl}} X_{kl} X_{ij}, \quad X_{ij}^2 = -1 \tag{4.3.16}$$

Braidless operations are therefore *projective representations* of adiabatic cycles. $\pm X_{ij}$ generate a non-Abelian double cover $\widetilde{\mathbb{Z}_2^{2N-1}}$, which is a non-split *central extention* of (4.3.13)

$$1 \to \mathbb{Z}_2 \to \widetilde{\mathbb{Z}_2^{2N-1}} \to \mathbb{Z}_2^{2N-1} \to 1 \tag{4.3.17}$$

4.3.3 Exchange operations

Exchange in $(3 + 1)$D is different from $(2 + 1)$D. Double exchange of Majorana fermions in $(3+1)$D can always be deformed away and left with a braidless operation

56

(fig.4.14). In other words, an exchange T_{ij} is half of a braidless operation $X_{ij} \propto \gamma_i\gamma_j = \exp(\pi\gamma_i\gamma_j/2)$. This suggests (up to a $U(1)$ phase) either

$$T_{ij} = \exp(\pi\gamma_i\gamma_j/4) \quad \text{or} \tag{4.3.18a}$$

$$T_{ij} = \exp(-\pi\gamma_i\gamma_j/4) \tag{4.3.18b}$$

The Ising exchange operator (4.3.18) can be illustrated explicitly by keeping track of the vector field $\vec{n}_t(\mathbf{r})$ in the bulk. Fig.4.15 are snapshots of an exchange between γ_1 and γ_2. $\vec{n}_t(\mathbf{r})$ points constantly at two distinct directions along the solid and dotted brown lines. Continuity of the vector field forbids crossing between a solid and a dotted line. A 2π twist always remains (see (t_4)) if the orientation of the hedgehogs are fixed. It can be untwisted by rotating a hedgehog hence resulting in a π-Berry phase at one Majorana fermion. The exchange operation in fig.4.15 therefore coincide with (4.3.18) so that $T_{12}\gamma_1 T_{12}^\dagger = -\gamma_2$ and $T_{12}\gamma_2 T_{12}^\dagger = \gamma_1$ for sequence (a), or $T_{12}\gamma_1 T_{12}^\dagger = \gamma_2$ and $T_{12}\gamma_2 T_{12}^\dagger = -\gamma_1$ for sequence (b). The sign that distinguish (4.3.18a) from (4.3.18b) is determined at step (t_3), which results in a 2π phase twist along either γ_2, γ_3 or γ_1, γ_4. The difference between (4.3.18a) and (4.3.18b) is a braidless operation, and therefore the sign difference is topologically protected by the inequivalent vector field evolution for (a) and (b).

The topology of the configuration space K_{2N}^{3D} is harder to describe when the Majorana fermions are allowed to move. It has been studied by Freedman *et.al.* [2011] and the fundamental group that topologically classified adiabtic cycles contains the
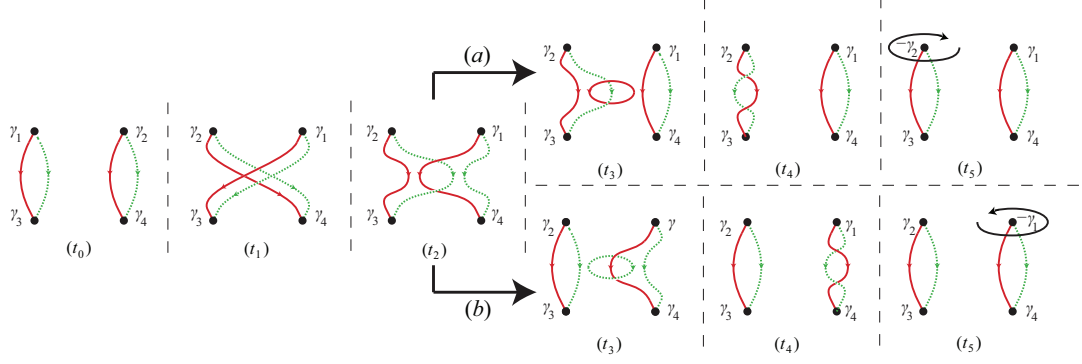
Figure 4.15: Evolution of $\vec{n}_t(\mathbf{r})$ in an exchange. Solid and dotted lines denote path of constant $\hat{n}(\mathbf{r})$. Orientation of hedgehogs are kept fixed throughtout t_0 to t_4. The 2π phase twist is elimnated at t_5 by rotating γ_2 resulting in a minus sign. (a) $T_{ij} = \exp(\pi\gamma_1\gamma_2/4)$; (b) $T_{ij} = \exp(-\pi\gamma_1\gamma_2/4)$.

essential piece

$$\pi_1(K_{2N}^{3D}) \supseteq E((\mathbb{Z}_2)^{2N} \rtimes S_{2N}) \tag{4.3.19}$$

the even part of $(\mathbb{Z}_2)^{2N} \rtimes S_{2N}$, called a *ribbon permutation group*. Note that the subgroup $E((\mathbb{Z}_2)^{2N}) = (\mathbb{Z}_2)^{2N-1}$ contains the braidless cycles (4.3.13), and the permutation group S_{2N} is responsible for exchanges. Exchange operations $\pm T_{ij}$ generate a *projective representation* of the ribbon permutation group. They satisfies the Yang-Baxter equation of the braid group

$$T_{ij}T_{jk}T_{ij} = T_{jk}T_{ij}T_{jk}, \quad \text{for distinct } i, j, k \tag{4.3.20}$$

with an extra constraint $T_{ij}^4 = -1$.[12]

---

[12] The Yang-Baxter equation holds for exchanges in the permutation group and the ribbon permutation group as well. The extra constraint is $\sigma_{ij}^2 = 1$ for permutation group, and $t_{ij}^4 = 1$ for ribbon permutation group. Both groups are finite.

Chapter 5

Electronic band theory and topological defects

The topological classification of electronic phases is a powerful tool in predicting the behavior of matter, especially along interfaces where gapless modes appear [Thouless *et.al.*, 1982; Halperin, 1982; Volovik and Yakovenko, 1989; Volovik, 1999; Read and Green, 2000; Kitaev, 2001; Volovik, 2003; Kane and Mele, 2005a; Kitaev, 2006; Fu, Kane and Mele, 2007; Volovik, 2009; Qi *et.al.*, 2009]. These robust mid-gap states, though exponentially localized at interface, cannot exist on their own or otherwise would violate the fermion doubling theorem [Nielssen and Ninomiya, 1983] and therefore must be connected with the bulk. They are irremovable by perturbation unless the energy gap or relevant symmetry of the bulk is destroyed. This accords with the definition 1.1.1 of bulk topological phases. The relationship between the topological class of the bulk system and the presence of robust mid-gap interface excitations, known as *bulk-boundary correspondence*, has been extensively

explored and verified in many electronic systems. An non-exhaustive list of them is given below.

TR breaking (class A)	2D:	QHE Haldane graphene model	chiral Dirac edge mode
TR symmetric (class AII)	2D:	QSHI	helical Dirac edge mode
	3D:	topological insulator	surface Dirac cone
PH symmetric (class D)	1D:	quantum Ising model Kitaev's SC chain	degenerate ordered ground state Majorana zero mode
	2D:	chiral p-wave SC Kitaev's honeycomb B-model	chiral Majorana edge mode
TR&PH symmetric (class DIII)	3D:	topological SC ^{3}He-B	surface Majorana cone

Table 5.1: Examples of bulk topological phases and boundary gapless modes. Abbreviations: TR = time reversal, PH = particle-hole, QHE = quantum Hall effect, QSHI = quantum spin Hall insulator, SC = superconductor.

A boundary gapless mode is a *topological soliton* associated with a domain wall separating a topological phase from a trivial one. A domain wall is the simplest kind of topological defect. It has *co-dimension* one in the sense that the boundary has one less dimensions than the bulk. There are other types of classical defects with different co-dimenions, such as vortices in the XY-model, half-vortices in smectics, dislocation in crystals, etc [Chaikin and Lubensky, 2000]. There is always a classical order parameter space $\mathcal{M}$ usually coming from symmetry breaking, and defects are classified by the homotopy group $\pi_D(\mathcal{M}) = [\mathbb{S}^D, \mathcal{M}]$, where $\mathbb{S}^D$ is some hypershere that wraps the defect.

In a topological band theory, there is in general no order parameter since topological insulating or superconducting phases are not described by spontaneous symmetry breaking. However, there is a topological description of how a band theory
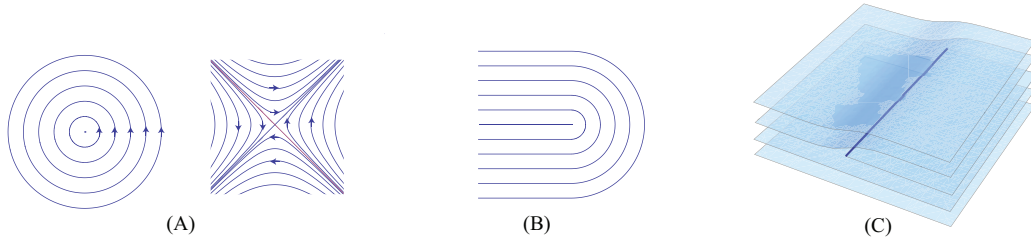
Figure 5.1: Classical topological defects. (A) ± 1 vortices in XY-model, order parameter $\theta \in U(1)$, $\pi_1(U(1)) = \mathbb{Z}$; (B) half vortex in smectics, order parameter $[\pm \hat{n}] = \mathbb{S}^1/\pm 1$, $\pi_1(\mathbb{S}^1/\pm 1) = \mathbb{Z}$; (C) edge dislocation in crystal, order parameter $[\mathbf{r}] \in \mathbb{R}^3/\mathbb{Z}^3$, $\pi_1(\mathbb{R}^3/\mathbb{Z}^3) = \mathbb{Z}^3$.

evolves around a defect [Teo and Kane, 2010a,b]. For instance, a topological invariant eq.(4.2.4) was used previously in section 4.2 to characterize a point defect in a superconductor and to indicate the presence of a Majorana bound state. This section will generalize this idea to topological defects in electronic band theory. It will focus on the topological classifications of spatially or temporally modulated band Hamiltonian, and the appearance of robust mid-gap defect excitation as a result of non-trivial modulations.

It will begin by providing a short review on the topological classification of single particle band Hamiltonians with full crystal translation symmetry. The 2+8 Atland Zirnbauer (AZ) symmetry classes [Cartan, 1926; Altland and Zirnbauer, 1997] and the periodic topological classification [Schnyder *et.al.*, 2008, 2009; Kitaev, 2009] will be introduced. A general framework describing topological defects in electronic band theory will be constructed in section 5.2. The classification in a symmetry class, whether it is 0, $\mathbb{Z}_2$ or $\mathbb{Z}$, will be shown to depend only on a single parameter $\delta = d - D$, called the defect dimension. Physical defects in low dimensions will be studied in the subsequent sections. This includes line defects, point defects and adiabatic pumps.

5.1 Topological band theory

A Bloch Hamiltonian of an insulator or a Bogoliubov de Gennes (BdG) Hamiltonian of a superconductor is a momentum parametrized matrix Hamiltonian $H(\mathbf{k})$ with a Fermi energy lying in a gapped energy spectrum. A local symmetry is represented by a unitary or antiunitary operator that commute or anticommute with the band Hamiltonian.[1] Local antiunitary symmetries involve complex conjugation $\mathcal{O}^{-1}e^{i\mathbf{k}\cdot\mathbf{r}}\mathcal{O} = e^{-i\mathbf{k}\cdot\mathbf{r}}$, and therefore operates on the Brillouin zone by the *involution* $\mathbf{k} \to -\mathbf{k}$.

1. **(Time reversal)** Any antiunitary operator Θ that commutes with the band Hamiltonian according to

$$\Theta^{-1}H(\mathbf{k})\Theta = H(-\mathbf{k}) \tag{5.1.1}$$

   qualifies as a time reversal (TR) symmetry.

2. **(Particle-hole)** Any antiunitary operator Ξ that anticommutes with the band Hamiltonian according to

$$\Xi^{-1}H(\mathbf{k})\Xi = -H(-\mathbf{k}) \tag{5.1.2}$$

   qualifies as a particle-hole (PH) symmetry.

---

[1]
(i) Only local symmetry will be considered here as space group symmetries are usually broken microscopically by defects.

(ii) Given a symmetry $\mathcal{O}$ such that $\mathcal{O}^{\dagger}H\mathcal{O} = e^{i\phi}H$, hermitianity of Hamiltonian requires $e^{i\phi} = \pm 1$.

(iii) There is no need considering a local unitary symmetry that commutes with H since all it does is decompose H into independent eigen-blocks.
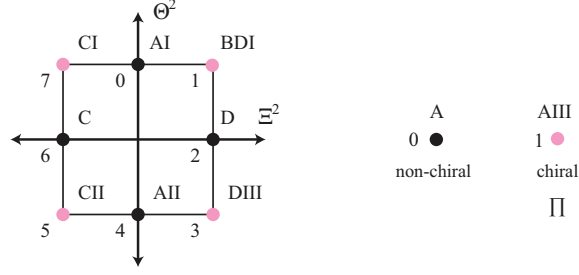
Figure 5.2: The 2+8 Altland Zirnbauer (AZ) symmetry classes of band Hamiltonians [Altland and Zirnbauer, 1997], labeled according to Cartan's classification of symmetric spaces [Cartan, 1926]. (Left) The 8 *real* symmetry classes that involve the antiunitary symmetries Θ (time reversal) and/or Ξ (particle-hole) are specified by the values of $\Theta^2 = \pm 1$ and $\Xi^2 = \pm 1$. They can be visualized on an eight hour "clock" [Teo and Kane, 2010b]. $s = 0, \ldots, 7 \pmod 8$ numerically labels the "hours". (Right) The 2 *complex* symmetry classes that does not involve antiunitary symmetries.

3. **(Chiral)** Any unitary operator Π anticommutes with the band Hamiltonian

$$\Pi^{-1} H(\mathbf{k}) \Pi = -H(\mathbf{k}) \tag{5.1.3}$$

qualifies as a chiral symmetry, such as the composition $\Xi\Theta$.

The square of any local symmetry operator $\mathcal{O}^2$ is a unitary operator that commutes with the Hamiltonian and therefore can be assumed to be constant by diagonalization. For antiunitary symmetries, $\mathcal{O}^{-1}(\mathcal{O}^2)\mathcal{O} = \mathcal{O}^2$ requires $\mathcal{O}^2 = \pm 1$; while for unitary symmetries, one can always assume $\mathcal{O}^2 = 1$ by redefining the operator by a phase factor. The presence or absence of the symmetries therefore form a total of 2+8 combinations, called symmetry classes due to Altland and Zirnbauer [1997].

The 8 *real* symmetry classes can be arranged on an eight hour "clock" (see fig.5.2) according to the signs of Θ^2 and Ξ^2 [Teo and Kane, 2010b]. The four chiral classes with both TR and PH symmetries are on the corners (e.g. class CI has $\Theta^2 = +1$ and

$\Xi^2 = -1$), while the other four non-chiral classes with either TR or PH symmetry are on the axes (e.g. class D has $\Xi^2 = +1$ but without TR symmetry). If both TR and PH are present, we may assume they commute $[\Theta, \Xi] = 0$. In general, $\mathcal{U} = \Xi^{-1}\Theta^{-1}\Xi\Theta$ is a unitary operator that commute with the Hamiltonian and therefore can be taken to be a constant. Hermitianity requires $\mathcal{U} = \pm 1$. Θ and Ξ can then be made to commute by multiplying a factor of i.

5.1.1 Periodic classifications

The set of stably topologically equivalent gapped band Hamiltonians within an AZ symmetry class s and in a certain dimension d form a *topological phase* (see definition 1.1.1 and the paragraph thereafter). Topological classification refers to the discrete collection of topological phases. The collection inherit an addition structure from direct sum of Hamiltonian matrices $H_1 \oplus H_2$.[2] The trivial phase corresponds to the additive identity, and the additive inverse of any Hamiltonian is the inversion between the conduction and valence bands about the Fermi energy. The collection of topological phases thus forms a discrete *group*, known as a topological K-group [Atiyah, 1994; Karoubi, 1978; Lawson and Michelsohn, 1989; Kitaev, 2009], denoted by $K_s(BZ)$ as it depends on the symmetries and the Brillouin zone.

The momentum space of a crystalline theory is a torus T^d. It contains many non-trivial lower dimensional cycles (sub-tori). These generates *weak* topologies as seen in the three weak indices of a 3D TR symmetric insulator [Fu, Kane and Mele, 2007].

---

[2]Given Hamiltonian matrices H_1 and H_2 with finite energy gaps, assuming they have the same Fermi energy,

$$H_1 \oplus H_2 = \begin{pmatrix} H_1 & 0 \\ 0 & H_2 \end{pmatrix} \tag{5.1.4}$$

is another gapped band Hamiltonian with the same Fermi energy and symmetries.

Weak topologies can be neglected by considering a continuum theory regularized at large momentum. The momentum space is *compactified* to a sphere $\mathbb{S}^d = \mathbb{R}^d \cup \{\infty\}$. Since there are no lower dimension cycles, only the *strong* topology remains. The group of strong topological classification will be denoted by $K(s; d) = K_s(\mathbb{S}^d)$.[3]

The topological classification was derived separately by Kitaev [2009] using Clifford algebra and Schnyder *et.al.* [2008, 2009] using random matrices. It is easiest to understand the stable classification by assuming the Hamiltonian has the Dirac-type structure[4] [Freedman *et.al.*, 2011; Stone, Chiu and Roy, 2011]

$$H(\mathbf{k}) = \hbar v_f \mathbf{k} \cdot \vec{\gamma} + m\Gamma \tag{5.1.6}$$

where $\gamma_1, \ldots, \gamma_d, \Gamma$ are hermitian anticommuting $2n \times 2n$ Dirac matrices that obey the TR and/or PH symmetry, and $n \to \infty$ for stability. The symmetry allowed mass terms $m\Gamma$ parametrize a manifold called a *classifying space* [Steenrod, 1999], denoted by $\mathcal{R}_{s-d}$ for the real classes and $\mathcal{C}_{s-d}$ for the complex ones. For instance, the mass term $m\Gamma(\mathbf{r})$ (c.f. vector field $\vec{n}(\mathbf{r})$) that describes the hedgehog around a Majorana fermion was shown to belong in U/O from eq.(4.2.12).

The stable classification is given by counting the number of connected components in classifying space.

$$K(s; d) = \pi_0(\mathcal{R}_{s-d}) \quad \text{or} \quad \pi_0(\mathcal{C}_{s-d}) \tag{5.1.7}$$

---

[3]The total group of classification over a toric Brillouin zone can be deduced combinatorically using the $C_r^d = \frac{d!}{r!(d-r)!}$ independent sub-tori that generates $H_r T^d$ [Kitaev, 2009].

$$K_s(T^d) = K_s(\mathbb{S}^d) \oplus K_s(\mathbb{S}^{d-1})^{\oplus C_{d-1}^d} \oplus \ldots \oplus K_s(\mathbb{S}^1)^{\oplus C_1^d} \oplus K_s(\mathbb{S}^0) \tag{5.1.5}$$

The addition structure however can be highly non-trivial and will not decompose into direct sume in general due to sub-tori intersection [Ran, 2010].

[4]This can be justified by a Morse theory argument given in our paper [Teo and Kane, 2010b].

for real or complex classes respectively. Bott periodicity theorem [Bott, 1956] tells

	Classifying spaces	π_0
$\mathcal{C}_0$	$\bigcup_{m>0} U(2m)/(U(m) \oplus U(m))$	$\mathbb{Z}$
$\mathcal{C}_1$	$(U(n) \times U(n))/U(n)$	0
$\mathcal{R}_0$	$\bigcup_{m>0} O(2m)/(O(m) \oplus O(m))$	$\mathbb{Z}$
$\mathcal{R}_1$	$(O(n) \times O(n))/O(n)$	$\mathbb{Z}_2$
$\mathcal{R}_2$	$O(2n)/U(n)$	$\mathbb{Z}_2$
$\mathcal{R}_3$	$U(2n)/Sp(n)$	0
$\mathcal{R}_4$	$\bigcup_{m>0} Sp(2m)/(Sp(m) \oplus Sp(m))$	$2\mathbb{Z}$
$\mathcal{R}_5$	$(Sp(n) \times Sp(n))/Sp(n)$	0
$\mathcal{R}_6$	$Sp(n)/U(n)$	0
$\mathcal{R}_7$	$U(n)/O(n)$	0

Table 5.2: Classifying spaces $\mathcal{R}_{s-d}$ and $\mathcal{C}_{s-d}$ for mass terms $m\Gamma$ of Dirac-type Hamiltonians and their connected components.

us the classifying spaces are related to each other by looping

$$\Omega \mathcal{R}_{s-d} = \mathrm{Map}(\mathbb{S}^1, \mathcal{R}_{s-d}) = \mathcal{R}_{s-d+1}, \quad \Omega^8 \mathcal{R}_{s-d} = \mathcal{R}_{s-d} \tag{5.1.8a}$$

$$\Omega \mathcal{C}_{s-d} = \mathrm{Map}(\mathbb{S}^1, \mathcal{C}_{s-d}) = \mathcal{C}_{s-d+1}, \quad \Omega^2 \mathcal{C}_{s-d} = \mathcal{C}_{s-d} \tag{5.1.8b}$$

This explains the group structure of (5.1.7) and the complex 2, real 8-periodicity in table 5.3.

	Symmetry				d							
s	AZ	Θ^2	Ξ^2	Π^2	0	1	2	3	4	5	6	7
0	A	0	0	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0
1	AIII	0	0	1	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$
0	AI	1	0	0	$\mathbb{Z}$	0	0	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$
1	BDI	1	1	1	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$
2	D	0	1	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0	$2\mathbb{Z}$	0
3	DIII	-1	1	1	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0	$2\mathbb{Z}$
4	AII	-1	0	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0
5	CII	-1	-1	1	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0
6	C	0	-1	0	0	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0
7	CI	1	-1	1	0	0	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$

Table 5.3: Periodic table for the stable topological classification $K(s;d)$ of crystalline insulators and superconductors [Schnyder *et.al.*, 2008, 2009; Kitaev, 2009]. The rows correspond to the different Altland Zirnbauer (AZ) symmetry classes s (mod 2 or 8), while the columns distinguish different dimensionalities d (mod 2 or 8).
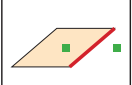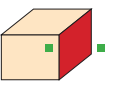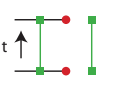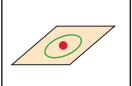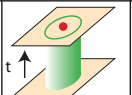
	$d=1$	$d=2$	$d=3$
$D=0$			
$D=1$			
$D=2$			

Figure 5.3: Topological defects characterized by a D parameter family of d dimensional Bloch-BdG Hamiltonians. Line defects correspond to $\delta = d - D = 2$, while point defects correspond to $\delta = d - D = 1$. Temporal cycles for point defects correspond to $\delta = d_d = 0$.

5.2 Defect classification

Topological defect in electronic band theory is desribed by a spatially and/or temporally modulated band Hamiltonian $H(\mathbf{k}, \mathbf{r})$ around the defect. For instance, a point defect in a superconductor was illustrated in eq.(4.2.8) (and (4.2.10)) with a position dependent vector field $\vec{n}(\mathbf{r})$ (resp. mass term $m\Gamma(\mathbf{r})$). Momentum $\mathbf{k}$ always lives in a crystalline Brillouin zone T^d or a continuum compactified space $\mathbb{S}^d = \mathbb{R}^d \cup \{\infty\}$. Position $\mathbf{r}$ lives away an ϵ-neighborhood off the defect in real space $\mathbb{R}^d - S_\epsilon$ so that a semi-classical description holds and $\mathbf{k}$ and $\mathbf{r}$ can be treated as commuting variables. This compliment of the ϵ-neighborhood deformation retract onto some submanifold (or cycles) Σ that wraps around the defect in real space. For example, a point defect is surround by a sphere $\Sigma = \mathbb{S}^{d-1}$ and a line defect is wrapped by a hypersphere $\Sigma = \mathbb{S}^{d-2}$. We are interested in local defect excitations and therefore can assume the defects to be linear subspaces: points, lines, planes, etc. They are always wrapped by hyperspheres $\Sigma = \mathbb{S}^D$. A time dependent adiabatic cycle depends on an extra temporal parameter and therefore $\Sigma = \mathbb{S}^{D-1} \times \mathbb{S}^1$.

We define the *defect dimension*

$$\delta = d - D \tag{5.2.1}$$

and we will show that the stable topological classification of a symmetry class depends only on this dimension parameter.

Time reversal and particle-hole are local symmetries and therefore are unbroken by any spatial or adiabatic temporal defect. Eq.(5.1.1), (5.1.2) and (5.1.3) change according to

$$\Theta^{-1}H(\mathbf{k},\mathbf{r})\Theta = H(-\mathbf{k},\mathbf{r}), \quad \Xi^{-1}H(\mathbf{k},\mathbf{r})\Xi = -H(-\mathbf{k},\mathbf{r}),$$

$$\text{and} \quad \Pi^{-1}H(\mathbf{k},\mathbf{r})\Pi = -H(\mathbf{k},\mathbf{r}) \tag{5.2.2a}$$

The stable classification of topological defect in band theory can then be formulated similar to that of the crystalline case. The collection of *defect topological classes*, i.e. stably equivalent classes of defect Hamiltonians, defines a additive group $K_s(BZ \times \mathbb{S}^D)$. The defect topology is carried by the highest dimension cycle in the base space $BZ \times \mathbb{S}^D$, analogous to the strong topology in crystalline band theory. This can be

singled out by *compactifying* the base space onto a sphere[5]

$$(k_1, \ldots, k_d; r_1, \ldots, r_D) \in \mathbb{S}^{d+D} = \mathbb{R}^{d+D} \cup \{\infty\} \tag{5.2.6}$$

and denote the defect classification by $K(s; d, D) = K_s(\mathbb{S}^{d+D})$.

By assuming a Dirac-type defect Hamiltonian

$$H(\mathbf{k}, \mathbf{r}) = \hbar v_f \mathbf{k} \cdot \vec{\gamma} + m\Gamma(\mathbf{r}) \tag{5.2.7}$$

the defect classification can be deduced by the homotopy groups of the classifying spaces

$$K_\mathbb{R}(s; d, D) = [\mathbb{S}^D, \mathcal{R}_{s-d}] = \pi_0(\Omega^D \mathcal{R}_{s-d}) = \pi_0(\mathcal{R}_{s-d+D}) \tag{5.2.8}$$

and similarly for the other 2 complex classes $K_\mathbb{C}(s; d, D) = \pi_0(\mathcal{C}_{s-d+D})$. Hence, this suggests that in a given symmetry class, the stable defect classification depends only on the dimension difference $\delta = d - D$ rather than the dimension d of the ambient material the defect lives in. For instance, this asserts the classification of a point defect of any symmetry would be the same in any dimension since a point defect has constant $\delta = d_D = 1$. Or in particular, line defects in 3D should have the

---

[5]The total group of defect classification is

$$K_s(T^d \times \mathbb{S}^D) = K(s; d, D) \oplus \sum_{r=0}^{d-1} K(s; r, D)^{\oplus C_r^d} \oplus \sum_{r=0}^{d} K(s; r, 0)^{\oplus C_r^d} \tag{5.2.3}$$

for time independent defects, or

$$K_s(T^d \times \mathbb{S}^{D-1} \times \mathbb{S}^1) = K(s; d, D) \quad \oplus \quad \sum_{r=0}^{d-1} K(s; r, D)^{\oplus C_r^d} \oplus \sum_{r=0}^{d} K(s; r, 0)^{\oplus C_r^d}$$

$$\oplus \quad \sum_{r=0}^{d} K(s; r, D-1)^{\oplus C_r^d} \oplus \sum_{r=0}^{d} K(s; r, 1)^{\oplus C_r^d} \tag{5.2.4}$$

for adiabatic cycles. A lower dimension weak topology can be removed from $H(\mathbf{k}, \mathbf{r})$ by subtraction

$$H_0(\mathbf{k}, \mathbf{r}) = H(\mathbf{k}, \mathbf{r}) \oplus -H'(\mathbf{k}', \mathbf{r}) \tag{5.2.5}$$

where $H'(\mathbf{k}', \mathbf{r})$ carries the weak topology for $\mathbf{k}'$ lives in some sub-torus $T^r \subseteq T^d$. The remaining Hamiltonian $H_0(\mathbf{k}, \mathbf{r})$ after removing all the weak topologies can then be compactified on a sphere $T^d \times \mathbb{S}^D \to \mathbb{S}^{d+D}$ since $H_0(k \to \infty, r \to \infty) \to$constant.

	Symmetry				$\delta = d - D$							
s	AZ	Θ^2	Ξ^2	Π^2	0	1	2	3	4	5	6	7
0	A	0	0	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0
1	AIII	0	0	1	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$
0	AI	1	0	0	$\mathbb{Z}$	0	0	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$
1	BDI	1	1	1	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$
2	D	0	1	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0	$2\mathbb{Z}$	0
3	DIII	-1	1	1	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0	$2\mathbb{Z}$
4	AII	-1	0	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0
5	CII	-1	-1	1	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0
6	C	0	-1	0	0	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0
7	CI	1	-1	1	0	0	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$

Table 5.4: Periodic table for the classification $K(s; d, D)$ of topological defects in insulators and superconductors. The original periodic table 5.3 is a particular case of this with $\delta = d - 0$.

same classification as 2D crystalline band theories, since the edge of a 2D system is also line defect with $\delta = 2$.

The classification according to classifying space of mass term of a Dirac-type Hamiltonian is flawed in the following two ways. (i) A defect band Hamiltonian in general is not of Dirac-type, and it is not immediately obvious how the topological classification $K(s; d, D)$ equates to the homotopy groups of symmetric spaces. (ii) A computational scheme of determining the defect topological class of a specific non-Dirac-type Hamiltonian is missing. These two issues will be addressed in the following subsections 5.2.1 and 5.2.2.

5.2.1   Bott periodicity

The topological defect classification $K(s; d, D)$ contains the following symmetries

$$K(s; d, D) \cong K(s + 1; d + 1, D) \cong K(s - 1; d, D + 1) \qquad (5.2.9)$$

For instances, it accounts for the diagonal recurrent pattern in the periodic table 5.4 and it also shows that the classification only depends on the defect dimension $\delta = d - D$ since $K(s; d, D) \cong K(s+1; d+1, D) \cong K(s; d+1, D+1)$. The complex 2 and real 8-periodicities are built-in because the symmetry label s is defined modulo 8.

$$K(s; d, D) \cong K(s; d+8, D) \cong K(s; d, D+8) \tag{5.2.10}$$

The isomorphisms in (5.2.9) can be constructed by simple explicit mappings between defect Hamiltonians $H_s(\mathbf{k}, \mathbf{r}) \mapsto H_{s+1}(\mathbf{k}, k', \mathbf{r})$ and $H_s(\mathbf{k}, \mathbf{r}) \mapsto H_{s-1}(\mathbf{k}, \mathbf{r}, r')$. By studying these, it will be shown that essentially all defect Hamiltonians can be deformed to have a Dirac-type structure (5.2.7). For simplicity, all Hamiltonians in this subsection is assumed to be *flattened* so that $\mathcal{H}^2 = \mathbb{1}$ and the Fermi energy lies at 0.[6]

The symmetry classes are divided into chiral and non-chiral ones depending on the presence or absence of chiral symmetry Π. A chiral Hamiltonian $\mathcal{H}_c$ can be map to a non-chiral one $\mathcal{H}_{nc}$ (fig.5.4 (left)) by adding an extra chiral symmetry breaking term

$$\mathcal{H}_{nc}(\mathbf{k}, \mathbf{r}, \theta) = \cos \theta \mathcal{H}_c(\mathbf{k}, \mathbf{r}) + \sin \theta \Pi \tag{5.2.11}$$

while a non-chiral Hamiltonian $\mathcal{H}_{nc}$ can be promoted into a chiral one $\mathcal{H}_c$ (fig.5.4 (right)) by doubling the number of bands

$$\mathcal{H}_c(\mathbf{k}, \mathbf{r}, \theta) = \cos \theta \mathcal{H}_{nc}(\mathbf{k}, \mathbf{r}) \otimes \tau_z + \sin \theta \mathbb{1} \otimes \tau_a, \tag{5.2.12}$$

---

[6]Band flattening can be achieved by deforming all the conduction (valence) bands to have energy $+1$ (resp. -1), and if necessary, add a number of trivial conduction or valence band so that there are equal number of bands below and above the Fermi energy. Given a Hamiltonian H, the flattened version $\mathcal{H}$ then loses all information on energetics, but preserves the topological phase. $\mathcal{H} = \mathbb{1} - 2P_v$, where P_v is the projection operator onto the valence band.
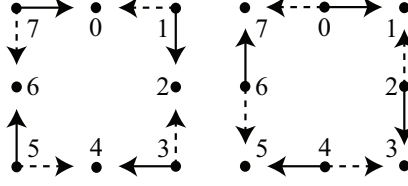
Figure 5.4: Hamiltonian mappings (5.2.11) and (5.2.12) are drawn on the left and right clocks respectively. Solid (dotted) arrows represents addition of one momentum (spatial) dimension.
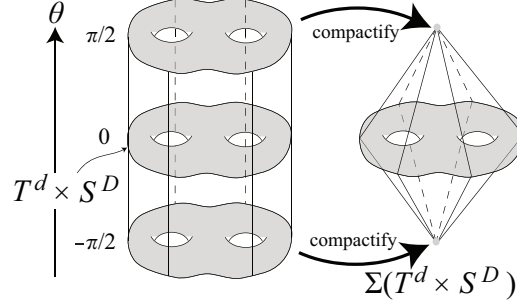
Figure 5.5: Supension $\Sigma(T^d \times \mathbb{S}^D)$. The top and bottom of the cylinder $(T^d \times \mathbb{S}^D) \times [-\pi/2, \pi/2]$ are identified into two points.

with chiral operator $\Pi = i\tau_z\tau_a$, for $a = x$ or y. The extra parameter θ in both of the equations can be either a momentum or spatial degree of freedom and lives in $[-\pi/2, \pi/2]$. Since the Hamiltonians are constant at $\theta = \pm\pi/2$, the parameter space for $(\mathbf{k}, \mathbf{r}, \theta)$ is *compactified* at the two points (see fig.5.5).

Depending on whether θ is a momentum parameter $\Theta, \Xi : \theta \to -\theta$ or spatial parameter $\Theta, \Xi : \theta \to \theta$, the extra term $\sin\theta\Pi$ in eq.(5.2.11) either destroys the TR or PH symmetry of the orginal Hamitonian $\mathcal{H}_c(\mathbf{k}, \mathbf{r})$; while the second term $\sin\theta\mathbb{1}\otimes\tau_a$ in eq.(5.2.12) must preserve the original symmetry of the original Hamltonian $\mathcal{H}_{nc}(\mathbf{k}, \mathbf{r})$. This fixes $\tau_a = \tau_x$ or τ_y and determines the symmetry class of $\mathcal{H}_c(\mathbf{k}, \mathbf{r}, \theta)$. It turns out that the addition of a momentum parameter θ always corresponds to a forward motion on the clock (solid arrow in fig.5.4) while a position parameter θ corresponds a backward motion (dotted arrows).

Eq.(5.2.11) and (5.2.12) can be shown, using an energy functional argument (appendix in [Teo and Kane, 2010b]), to be isomorphisms (invertible maps preserving group addition structure) in K-theory

$$K(s; d, D) \xrightarrow{\cong} K(s + 1; d + 1, D) \quad \text{and} \quad K(s; d, D) \xrightarrow{\cong} K(s - 1; d, D - 1) \quad (5.2.13)$$

In particular this means that any continuum Hamiltonian $\mathcal{H}_s$ is stable equivalent to either of the following form

$$\mathcal{H}_s(k_1, \ldots, k_d, \mathbf{r}) \simeq \mathcal{H}_{s-1}(k_1, \ldots, k_{d-1}, \mathbf{r}) + k_d \Pi + O(k_d^2) \qquad (5.2.14)$$

$$\mathcal{H}_s(k_1, \ldots, k_d, \mathbf{r}) \simeq \mathcal{H}_{s-1}(k_1, \ldots, k_{d-1}, \mathbf{r}) \otimes \tau_z + k_d \tau_a + O(k_d^2) \quad (5.2.15)$$

depending whether the symmetry class s is chiral or non-chiral. Here the parameter space is compactified as $(\mathbf{k}, \mathbf{r}) \in \mathbb{S}^{d+D}$ and the trigonometric functions can be Taylor expand around the origin to be $\cos \theta = 1 + O(k_d^2)$ and $\sin \theta = k_d + O(k_d^3)$. Given a sufficient number of bands, this identification process can be iterated so that

$$\begin{aligned} \mathcal{H}_s(k_1, \ldots, k_d, \mathbf{r}) &\simeq \mathcal{H}_{s-1}(k_1, \ldots, k_{d-1}, \mathbf{r}) + k_d \Pi + O(k_d^2) \\ &\simeq \mathcal{H}_{s-2}(k_1, \ldots, k_{d-2}, \mathbf{r}) \otimes \tau_z + k_{d-1} \tau_a + k_d \Pi \\ &\simeq \ldots \\ &\simeq \mathcal{H}_{s-d}(\mathbf{r}) \otimes \Gamma + k_1 \gamma_1 + \ldots + k_d \gamma \qquad (5.2.16) \end{aligned}$$

a Dirac-type Hamiltonian. The mass term $\mathcal{H}_{s-d}(\mathbf{r}) \otimes \Gamma$ can then be classified according to the homotopy groups of classifying space $\pi_D(C_{s-d})$.

5.2.2 Integral formulae

The defect topological class of most Hamiltonians can be determined using certain integral formulae, known as topological invariants. Examples include the Chern invariant of an integer quantum Hall state eq.(1.1.2), the $\mathbb{Z}_2$-invariant for topological insulators in 2 and 3D eq.(1.1.14, 1.1.16, 1.1.18), and the Chern-Simons invariant for Majorana bound state eq.(4.2.4). In general, the topological invariant can mostly be universally written down according to the *dimension hierachy* [Qi, Hughes and Zhang, 2008; Ryu *et.al.*, 2010] along rows in the periodic table 5.4

$$
\underset{s-d+D\equiv}{\overset{\text{2nd descendant}}{\mathbb{Z}_2}} \xleftarrow[\cong]{i^*} \underset{-1}{\overset{\text{1st descendant}}{\mathbb{Z}_2}} \xleftarrow[\text{mod 2}]{i^*} \underset{0}{\mathbb{Z}} \tag{5.2.17}
$$

Here i^* restricts a model on a sphere $(\mathbf{k}, \mathbf{r}) \in \mathbb{S}^{d+D}$ onto its equator $\mathbb{S}^{(d-1)+D}$.

For simplicity, all Hamiltonians $H(\mathbf{k}, \mathbf{r})$ are assumed to be compactified so that $(\mathbf{k}, \mathbf{r}) \in \mathbb{S}^{d+D}$ (see footnote below eq.(5.2.6)). The invariants are constructed by the valence states $|u_m(\mathbf{k}, \mathbf{r})\rangle$ and the Berry connection $\mathcal{A}(\mathbf{k}, \mathbf{r})$ eq.(4.2.6). Chiral and non-chiral models (resp. odd and even s) have different types of invariants. For chiral classes, it is most convenient to assume $\Pi = \tau_z$ (by a gauge transformation in general) so that the *flattened* Hamiltonian takes the off-block diagonal form

$$
\mathcal{H}_c(\mathbf{k}, \mathbf{r}) = \begin{pmatrix} 0 & q(\mathbf{k}, \mathbf{r}) \\ q(\mathbf{k}, \mathbf{r})^\dagger & 0 \end{pmatrix}, \quad q(\mathbf{k}, \mathbf{r}) \in U(n) \tag{5.2.18}
$$

For non-chiral classes, the topological invariants are often composed of the gauge

covariant Berry curvature

$$\mathcal{F} = d\mathcal{A} + \mathcal{A} \wedge \mathcal{A} \tag{5.2.19}$$

	2nd $\mathbb{Z}_2$ descendant	1st $\mathbb{Z}_2$ descendant	$\mathbb{Z}$ or $2\mathbb{Z}$
non-chiral	Fu-Kane	Chern-Simons	Chern character
chiral	Chern-Simons	Fu-Kane	winding number

Table 5.5: Universal topological invariants.

1. **($\mathbb{Z}$-invariants)** This applies to $s - d + D \equiv 0$ (mod 4). For non-chiral classes, the integral invariant is given by the Chern character that classifies the *vector bundle* of valence states

$$\text{NC:} \qquad n = \frac{1}{(\Delta/2)!} \left(\frac{i}{2\pi}\right)^{\Delta/2} \int_{\mathbb{S}^{d+D}} \text{Tr}[\underbrace{\mathcal{F} \wedge \ldots \wedge \mathcal{F}}_{\Delta/2}] \tag{5.2.20}$$

while for chiral classes, the invariant is given by the winding number that characterized the stable homotopty of $g(\mathbf{k}, \mathbf{r})$ in $\pi_{d+D}(U(n))$

$$\text{C:} \qquad n = \frac{\left(\frac{\Delta-1}{2}\right)!}{\Delta!(2\pi i)^{\frac{\Delta+1}{2}}} \int_{\mathbb{S}^{d+D}} \text{Tr}[\underbrace{(qdq^\dagger) \wedge \ldots \wedge (qdq^\dagger)}_{\Delta}] \tag{5.2.21}$$

where $\Delta = d + D$ is the total dimension and n is an integer for both of them. n must be even for $s - d + D \equiv 4$ (mod 8). The two invariants (5.2.21) and (5.2.20) are interchangeable by the isomorphisms (5.2.11) and (5.2.12) that relates chiral and non-chiral classes.

2. **(1st $\mathbb{Z}_2$ descendant)** This applies to $s - d + D \equiv -1$ (mod 8). For non-chiral

classes, dimension hierachy (5.2.17) imposes a Chern-Simons invariant

$$\text{NC:} \qquad \nu \equiv \frac{2}{\left(\frac{\Delta+1}{2}\right)!} \left(\frac{i}{2\pi}\right)^{\frac{\Delta+1}{2}} \int_{\mathbb{S}^{d+D}} \mathcal{Q}_{\Delta} \qquad \text{mod } 2 \qquad (5.2.22)$$

where $\mathcal{Q}_{\Delta}$ is the Chern-Simons Δ-form [Chern and Simons, 1974; Nakahara, 1990]

$$\mathcal{Q}_{\Delta} = \frac{\Delta+1}{2} \int_0^1 dt \text{Tr} \left[\mathcal{A} \left(t d\mathcal{A} + t^2 \mathcal{A} \wedge \mathcal{A} \right)^{\frac{\Delta-1}{2}} \right] \qquad (5.2.23)$$

The defect classification of a chiral class s is identical to that of the 2nd descendant in the non-chiral class $s+1$. This is established from the dimension hierachy and the isomorphism (5.2.11).

$$\begin{array}{ccccc} \mathbb{Z}_2 & \xleftarrow{i^*} & \mathbb{Z}_2 & \xleftarrow{i^*} & \mathbb{Z} \\ \| & & \cong & & \cong \\ \mathbb{Z}_2 & \xleftarrow{i^*} & \mathbb{Z}_2 & \xleftarrow{i^*} & \mathbb{Z} \end{array} \qquad (5.2.24)$$

where the top row is the chiral class s and the bottom one is the non-chiral $s+1$. This is essentially treating the chiral Hamiltonian as a non-chiral one by *forgetting* either TR or PH symmetry. The subsequent non-chiral model is characterized by the Fu-Kane invariant (5.2.25) below.

3. **(2nd $\mathbb{Z}_2$ descendant)** This applies to $s - d + D \equiv -2 \pmod 8$. The integral invariants for both non-chiral and chiral cases involve the Berry connection $\mathcal{A}$. It has to be constructed from valence states $|u_m(\mathbf{k}, \mathbf{r})\rangle$ that obey a certain gauge constraint, or otherwise the invariant would be vacuous. For non-chiral classes, the invariant is given by a formula that resemblant the Fu-Kane invariant

(1.1.16) for 2D topological insulators

$$\text{NC:} \qquad \tilde{\nu} \equiv \frac{i^{\Delta/2}}{(\Delta/2)!(2\pi)^{\Delta/2}} \left[\int_{\mathbb{S}_{1/2}^{d+D}} \text{Tr}(\mathcal{F}^{\Delta/2}) - \oint_{\partial\mathbb{S}_{1/2}^{d+D}} \mathcal{Q}_{\Delta-1} \right] \qquad \text{mod } 2$$

$$(5.2.25)$$

where $\mathbb{S}_{1/2}^{d+D}$ is a hemisphere, and $\partial\mathbb{S}_{1/2}^{d+D} = \mathbb{S}^{(d-1)+D}$ is its equatorial boundary that is closed under the involution $(\mathbf{k}, \mathbf{r}) \to (-\mathbf{k}, \mathbf{r})$. The Chern-Simons $(\Delta-1)$-form is constructed with a gauge that satisfies constraint (5.2.27) or (5.2.30) along the equator. For chiral classes, they are classified by a Chern-Simons invariant that differs (5.2.22) by a factor of 2,

$$\text{C:} \qquad \tilde{\nu} \equiv \frac{1}{\left(\frac{\Delta+1}{2}\right)!} \left(\frac{i}{2\pi}\right)^{\frac{\Delta+1}{2}} \int_{\mathbb{S}^{d+D}} \mathcal{Q}_\Delta \qquad \text{mod } 2 \qquad (5.2.26)$$

with the gauge constraint taken over the base space $\mathbb{S}^{d+D}$. The gauge constraint depends on the symmetry class. For class AI, DIII, AII and CI ($s = 0, 3, 4, 7$), a time reversal constraint should be taken

$$w_{mn}(\mathbf{k}, \mathbf{r}) = \langle u_m(\mathbf{k}, \mathbf{r}) | \Theta u_n(-\mathbf{k}, \mathbf{r}) \rangle \equiv \text{constant} \qquad (5.2.27)$$

while for class BDI, D, CII and C ($s = 1, 2, 5, 6$), a more complicated and non-deterministic particle-hole constraint should be used. Given the conduction and valence frame u_c and u_v, one may combine them to form a unitary matrix

$$G_{\mathbf{k},\mathbf{r}} = \begin{pmatrix} | & | \\ u_v(\mathbf{k}, \mathbf{r}) & u_c(\mathbf{k}, \mathbf{r}) \\ | & | \end{pmatrix} \qquad (5.2.28)$$

77

The orthogonality of conduction and valence band states implies that

$$G_{\mathbf{k},\mathbf{r}}^{\dagger} \Xi G_{-\mathbf{k},\mathbf{r}} = 0 \qquad (5.2.29)$$

The particle-hole gauge constraint requires the matrix G to be homotopically trivial

$$G_{\mathbf{k},\mathbf{r}} \simeq \text{constant} \qquad (5.2.30)$$

i.e. $G_{\mathbf{k},\mathbf{r}}^{\dagger}$ should be able to deformed continuously into a constant matrix while maintaining (5.2.29) along the deformation.

5.3 Physical defects in low dimensions

The topology of defect Bloch-BdG Hamiltonian $H(\mathbf{k}, \mathbf{r})$ is physically revealed by robust low energy defect excitations that violates the fermion doubling theorem [Nielssen and Ninomiya, 1983]. These include boundary excitations such as the surface Dirac cone of a topological insulator in section 2 or the helical edge mode of a quantum spin Hall insulator in section 3. Majorana zero modes in chapter 4 are low energy excitations associated to a point defect, and braidless operations are consequence of non-trivial adiabatic cycles.

Various topological defect excitations associated to line, point defects and adiabatic cycles will be investigated below. Explicit examples will be introduced as theoretical proposals or be revisited as illustrations of the generalized bulk-boundary correspondence. In particular, the semi-classical approach of defect band Hamil-

Symmetry	Topological classification	1D Gapless Fermion modes
A	$\mathbb{Z}$	Chiral Dirac
D	$\mathbb{Z}$	Chiral Majorana
DIII	$\mathbb{Z}_2$	Helical Majorana
AII	$\mathbb{Z}_2$	Helical Dirac
C	$2\mathbb{Z}$	Chiral Dirac

Table 5.6: Symmetry classes that support topologically non trivial line defects and their associated protected gapless modes.

tonians $H(\mathbf{k}, \mathbf{r})$ will be applied to understand heterostructures and lattice defects. The topological classification and characterization of defect band Hamiltonians then becomes a powerful tool in predicting the existence of topological defect modes.

5.3.1 Line defects

A line defect has dimension $\delta = d - D = 2$. $(\mathbf{k}, \mathbf{r})$ lives in the base space $T^d \times \mathbb{S}^{d-2}$ with $\mathbb{S}^{d-2}$ wrapping the defect line in real d-space. Robust low energy line excitations are characterized according to the non-zero entries along the $\delta = 2$ column of the periodic table 5.4. Line defects in 2 dimensions can be realized as edges of a bulk material and have already been extensively studied. It will therefore be more interesting to consider line defects in 3 dimensions instead. In fact, it may even be technological more feasible to manufacture a 3 dimensional structure than a 2 dimensional one.
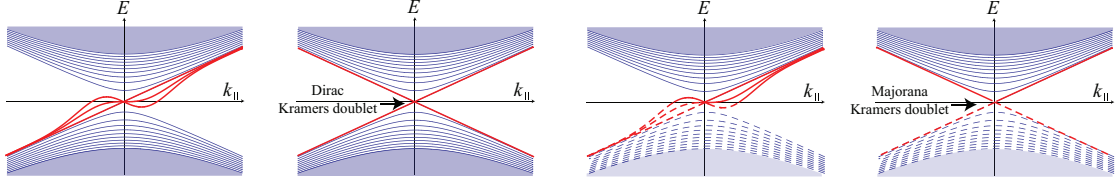
Figure 5.6: 1D Gapless excitations. (Left to right) chiral Dirac mode ($n = 3$), helical Dirac mode, chiral Majorana mode ($n = 3$), helical Majorana mode. $k_\parallel$ is momentum along the line defect.

Heterostructures

A line defect is bounded in 3D at the interface between three different bulk gapped electronic materials (see fig.5.7) such as trivial insulators (air or vacuum) - I, topological insulators - TI, (anti)ferromagnets - AM/FM, superconductors - SC with different pairing phases. In the vicinity of the line interface, the defect can be modeled by a semi-classical Dirac type Hamiltonian

$$H(\mathbf{k}, s) = \hbar v_f \mathbf{k} \cdot \vec{\gamma} + m\Gamma(s) \tag{5.3.1}$$

on a circle surrounding the line. The mass term differs in the three materials, and would be discontinuous across sharp surface interfaces. However, by requiring the absence of phase transition across any pair of materials, the surface interfaces must remain gapped and therefore can be smoothened out (c.f. the smooth interface across the TI-SC-I junction in fig.4.8). The defect topology is described by the winding of the mass term $m\Gamma(s)$ around the line junction and corresponds to gapless 1D excitation. It must be emphasized that *gapped* surface interfaces between pairs are necessary conditions in hosting a topological line excitation. Or otherwise the gapless mode could escape along the surfaces, and in fact the geometry could no longer be considered as a line defect since the mass term $m\Gamma(s)$ would be singular
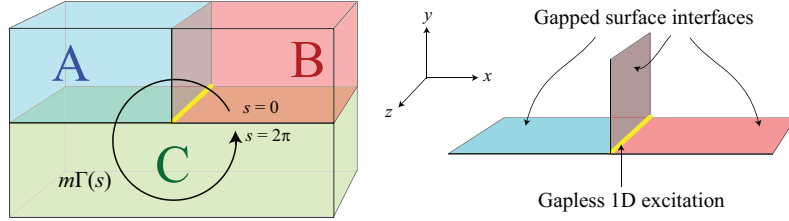
Figure 5.7: Line defect at a heterostructure. A, B and C are different bulk gapped materials put together so that there is no gapless surface modes along interfaces of any pairs. The mass term $m\Gamma(s)$ wraps non-trivially around the line interface corresponding to a gapless 1D excitation.
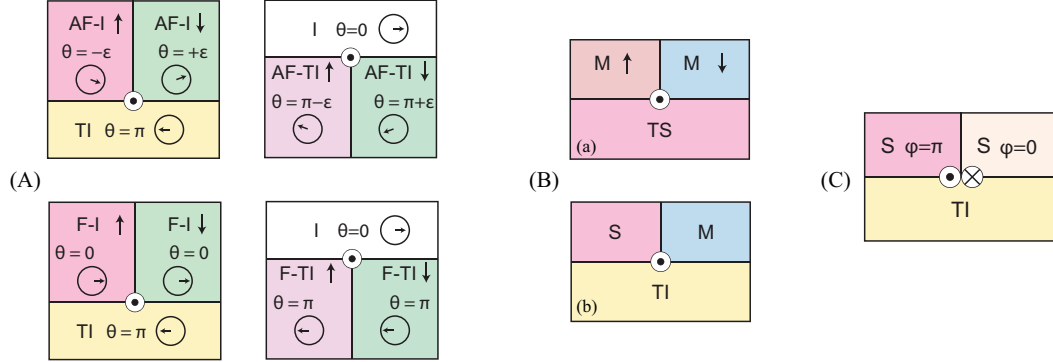
Figure 5.8: Heterostructures' topologies and excitations. Abbreviations: AF = antiferromagnetic, F = ferromagnetic, I = trivial insulator, TI = topological insulator, S = superconductor, TS = TR-symmetric topological superconductor, M = magnet. (A) Chiral Dirac mode protected by winding of the magnetoelectric θ-angle. (B) Chiral Majorana mode protected by a second Chern number. (C) Helical Majorana mode protected by a Fu-Kane invariant.

across a phase transition.

A few theoretical proposals on heterostructures are given below. Although one would have to overcome the technological difficulty of manufacturing smooth surface interfaces with an appropriate gap, the 1D gapless excitations are otherwise robust and guaranteed by topology. In a realistic system, the topology is characterized by the $\mathbb{Z}$ or $\mathbb{Z}_2$ invariants layed down previously in section 5.2.2. And these identify with the winding of $m\Gamma(s)$ in the phenomenological Dirac-type model.

A) [(**Anti)ferromagnetic topological insulator**] The geometries are specified by fig.5.8(A), and the charge conservation symmetry is specified by the complex

AZ class A. They can be approximated by the Dirac Hamiltonians close to the line interface

$$H(\mathbf{k}, s) = \hbar v_f \mu_x \vec{\sigma} \cdot \mathbf{k} + m(s)\mu_z + \begin{cases} h_{af}(s)\mu_y \\ \vec{h}_f(s) \cdot \vec{\sigma} \end{cases} \tag{5.3.2}$$

Here $m(s)$ describes the topological to trivial band inversion as it changes signs along the vertical axis. h_{af} and $\vec{h}_f$ are TR-breaking (anti)ferromagnetic parameters, distinguished by behavior under inversion μ_z. It is easiest to understand in the AF-case, where the antiferromagnet orientation switches across the horizontal axies and $h_{af}(s)$ simply changes signs. Phenomenologically,

$$m\Gamma(s) \sim (m_0 \sin s)\mu_z + (h_0 \cos s)\mu_y \tag{5.3.3}$$

which is a 360° rotation in the μ degree of freedom. This winding number can be identified with the circulation of the magnetoelectric θ-angle (1.1.18) [Qi, Hughes and Zhang, 2008]. In general, the number of 1D chiral Dirac modes is given by how many times the θ-angle winds around the line junction.

$$n = \frac{1}{2\pi} \int_{\mathbb{S}^1} d\theta(s) = \frac{-1}{4\pi} \int_{BZ \times \mathbb{S}^1} \text{Tr}[\mathcal{F} \wedge \mathcal{F}] \tag{5.3.4}$$

The heterostructure therefore supports a local unidirectional transport channel of charge in three dimensions with $e^2/2h$ conductance that is robust against backscattering and localization.

B) [**Magnetic topological superconductor**] The heterostructure geometries are given by fig.5.8(B)[7] with the superconductor breaking charge conservation. The

---

[7]In (a), the 3D bulk TR-preserving superconductor is assumed to carry $n = 1$ topology in class DIII. This is to

defects are described by the Dirac-type BdG Hamiltonians in class D

$$H(\mathbf{k}, s) = \hbar v_f \tau_z \vec{\sigma} \cdot \mathbf{k} + \Delta(s)\tau_x + h(s)\tau_y \tag{5.3.5a}$$

$$H(\mathbf{k}, s) = \tau_z \left[\hbar v_f \mu_x \vec{\sigma} \cdot \mathbf{k} + m(s)\mu_z \right] + \Delta(s)\tau_x + h(s)\mu_y \tag{5.3.5b}$$

respectively for (a) and (b) in fig.5.8(B).[8] The first Hamiltonian has the same

form of the antiferromagnetic Hamiltonian (5.3.2), and therefore carries the

same unit winding $m\Gamma(s) = (\Delta_0 \sin s)\tau_x + (h_0 \cos s)\tau_y$ that guarantees a single

chiral Majorana mode along the line interface. The three mass m, Δ, h terms

in the second Hamiltonian are controlled by the I-TI transition, supconducting

and magnetic strength. $m(s)$ changes signs along the vertical axis. Supercon-

ductivity dominates $\Delta \gg h$ for $x < 0$ and magnetism dominates $\Delta \ll h$ for

$x > 0$. The Hamiltonian can be decomposed $H = H_+ \oplus H_-$ according to the

good quantum number $\tau_x \mu_y = \pm 1$, and only one of the sector carries non-trivial

topology.

$$H_-(\mathbf{k}, s) = \hbar v_f \tilde{\tau}_z \vec{\sigma} \cdot \mathbf{k} + m(s)\tilde{\tau}_x + (\Delta(s) - h(s))\tilde{\tau}_y \tag{5.3.6}$$

Again $(m, \Delta - h)$ winds 360° around the line junction. Thus both heterostructure

(a) and (b) support local unidirectional transport channel of neutral Majorana

fermions and provide possible ways of manipulating the exotic excitation in 3D.

C) [**Josephson junction over topological insulator**] The details are specified

in fig.5.8(C) [Fu, Kane and Mele, 2007]. Time reversal symmetry is preserved

when the superconducting phase is tuned at exactly $\varphi = \pm\pi$. It can be modeled

---

guarantee a single chiral Majorana channel.

[8]We assume antiferromagnetic coupling $h(s)$ for simplicity.
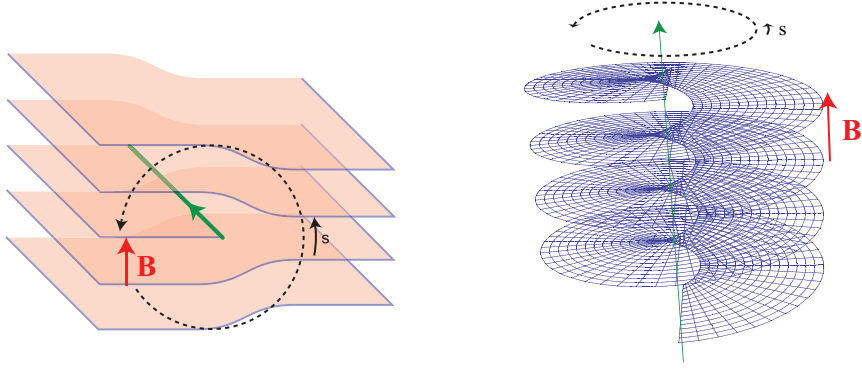
Figure 5.9: Edge and screw dislocations with their Burger's vector (red) perpendicular and parallel to the line defect (green). A general dislocation can be a linear combination of the two, and the bulk crystal do not have to be stratified in any particular direction.

by the minimal BdG Dirac Hamiltonian in class DIII

$$H(\mathbf{k}, s) = \tau_z\left[\hbar v_f \mu_x \vec{\sigma} \cdot \mathbf{k} + m(s)\mu_z\right] + \Delta(s)\tau_x \qquad (5.3.7)$$

The phenomenological mass term $m\Gamma(s) = (m_0 \sin s)\tau_z\mu_z + (\Delta_0 \cos s)\tau_x$ has unit winding.[9] The topology can be more generally characterized by the $\mathbb{Z}_2$ Fu-Kane invariant (5.2.25). The non-triviality of that ensures the existence of helical Majorana mode.

Dislocations

Dislocations are classical topological line defects in 3D crystals. Each is characterized by a *Burger's vector*, a Bravais lattice vector $\mathbf{B} = m_1\mathbf{a}_1 + m_2\mathbf{a}_2 + m_3\mathbf{a}_3$, that measures the displacement of a density wave along a cycle around the line defect.[10] A dislocation in an electronic material with non-trivial topology could carry a gapless 1D excitation.

---

[9] To be precise, $m\Gamma(s)$ lives in the classifying space $\mathcal{R}_0 = O(2m)/O(m) \times O(m)$, the real Grassmann manifold. The winding is defined modulo 2 as $\pi_1(\mathcal{R}_0) = \mathbb{Z}_2$.

[10] A density wave has order parameter in $\mathbb{R}^3/\mathcal{L}$, $\mathcal{L} \cong \mathbb{Z}^3$ is the Bravais lattice. $\mathbf{B} \in \pi_1(\mathbb{R}^3/\mathcal{L}) = \mathcal{L}$.

A line defect in a TR-breaking 3D material is characterized by a second Chern number (5.3.4). For a dislocation in a three dimensional quantum Hall state, this invariant can be simplified into a dot product

$$n = \frac{1}{2\pi} \mathbf{B} \cdot \mathbf{G}_c \tag{5.3.8}$$

where $\mathbf{G}_c$ is the reciprocal lattice vector that characterized the 3D quantum Hall state

$$\mathbf{G}_c = \frac{1}{2\pi} \int_{BZ} d\mathbf{k} \wedge \mathrm{Tr}[\mathcal{F}_0] \tag{5.3.9}$$

$\mathcal{F}_0(\mathbf{k})$ is the Berry curvature for the crystalline bulk material. The integral invariant n determines the number of chiral Dirac excitation running along the dislocation.

Dislocations in weak topological insulators were studied by Ran [Zhang and Vishwanath]. This could be understood as a defect band theory in class AII and characherized using the Fu-Kane $\mathbb{Z}_2$ invariant from (5.2.25).

$$\tilde{\nu} \equiv \frac{1}{8\pi^2} \left(\int_{\frac{1}{2}T^3 \times \mathbb{S}^2} \mathrm{Tr}[\mathcal{F} \wedge \mathcal{F}] - \int_{\partial \frac{1}{2}T^3 \times \mathbb{S}^2} \mathcal{Q}_3 \right) \qquad \mathrm{mod}\ 2 \tag{5.3.10}$$

It can be simplified into a dot product similar to (5.3.8)

$$\tilde{\nu} \equiv \frac{1}{2\pi} \mathbf{B} \cdot \mathbf{G}_\nu \qquad \mathrm{mod}\ 2 \tag{5.3.11}$$

where $\mathbf{G}_\nu$ is the $\mathbb{Z}_2$-valued reciprocal lattice vector that characterized weak topolgies

$$\mathbf{G}_\nu = \int_{\frac{1}{2}T^3} \mathrm{Tr}[\mathcal{F}_0] \wedge d\mathbf{k} - \int_{\partial \frac{1}{2}T^3} \mathrm{Tr}[\mathcal{A}_0] \wedge d\mathbf{k} \tag{5.3.12}$$

$\mathcal{A}_0(\mathbf{k})$ and $\mathcal{F}_0(\mathbf{k})$ are the Berry connection and curvature for the crystalline bulk. The $\mathbb{Z}_2$-invariant indicates the existence or absence of a helical Dirac mode along the line dislocation.

Other 1D gapless excitations are associated with dislocations in different crystalline electronic materials. Chiral Majorana mode could appear in a 3D chiral p-wave superconductor such as Sr_2RuO_4, and helical Majorana mode could exist in a *weak* TR-symmetric topological superconductor.

5.3.2 Point defects

A point defect has dimension $\delta = d - D = 1$. $(\mathbf{k}, \mathbf{r})$ lives in the base space $T^d \times \mathbb{S}^{d-1}$ with $\mathbb{S}^{d-1}$ wrapping the point defect in real d-space. Robust zero energy bound states are characterized according to the non-zero entries along the $\delta = 1$ column of the periodic table 5.7. Chiral Dirac terminal mode of the Su-Schrieffer-Heeger model [1980] and the Jackiw-Rebbi model [1976], Majorana zero modes at the ends of the Kitaev's superconducting chain [Kitaev, 2001], at vortex cores of a chiral p-wave superconductor [Volovik, 1999; Read and Green, 2000] and the Jackiw-Rossi model [1981] are examples of robust bound states associated with a point defect. The system always has a chiral or particle-hole symmetry that anticommute with the Bloch-BdG Hamiltonian so that a bound state with $\Pi|\psi_0\rangle = \pm|\psi_0\rangle$ or $\Xi|\psi_0\rangle = |\psi_0\rangle$ is pinned exactly at zero energy. A non-degenerate zero energy bound state is irremovable unless the symmetry is broken or the bulk gap closes.

Symmetry	Topological classification	$E = 0$ Bound States
AIII	$\mathbb{Z}$	Chiral Dirac
BDI	$\mathbb{Z}$	Chiral Majorana
D	$\mathbb{Z}_2$	Majorana
DIII	$\mathbb{Z}_2$	Majorana Kramers doublet (= Dirac)
CII	$2\mathbb{Z}$	Chiral Majorana Kramers doublet (= Chiral Dirac)

Table 5.7: Symmetry classes supporting non trivial point topological defects and their associated $E = 0$ modes.
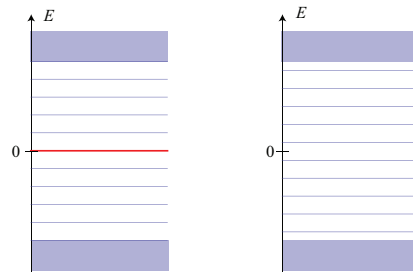
Figure 5.10: Particle-hole symmetric bound states of a point defect. (Left) Zero energy bound state (red) at a topological point defect. (Right) Removable $(+E, -E)$-pair of bound states at a trivial point defect.

Chiral zero mode   Given a chiral symmetry in the system, the zero energy bound states decompose according to their *chirality*, i.e. eigenvalues of the chiral operator $\pi = \pm 1$. Setting $\Pi = \tau_z$, the defect Hamiltonian operator[11] takes the off-block diagonal form

$$\hat{H} = \hat{h} + \hat{h}^\dagger, \quad \hat{h} = \hat{q} \otimes \tau_+ \tag{5.3.13}$$

where $\tau_\pm = (\tau_x \pm i\tau_y)/2$, $\hat{q}$ is an invertible operator. $\hat{h}$ is known mathematically as a Dirac operator [Nakahara, 1990; Berline, Getzler and Vergne, 2004]. It sends a negative chiral state to a positive one while annihilating all positive chiral states. The *chirality* of zero energy bound states are defined as an *analytic index*

$$\mathrm{ind}(\hat{H}) = \dim \ker(\hat{h}^\dagger) - \dim \ker(\hat{h}) \tag{5.3.14}$$

which is the difference between the number of positive and negative chiral zero modes.[12] Since a pair of zero modes could splits into $\pm E$ pairs only when they have opposite chirality, the analytic index (chirality) is robust against any perturbations on the Hamiltonian operator as long as the chiral symmetry and its ellipticity (bulk gap) is maintained.

The topology of the defect is described by the semi-classical theory $H(\mathbf{k}, \mathbf{r})$. The *topological index* is given by the winding number (5.2.21). It is an integral invariant

---

[11]Given a semi-classical band Hamiltonian $H(\mathbf{k}, \mathbf{r})$ away from a point singularity, a defect Hamiltonian operator $\hat{H}$ can be defined essentially by replacing $\mathbf{k} \leftrightarrow -i\nabla$. This can be made mathematically by introducing *pseudo-differential operators* using Fourier transform, and the ellipticity of the operator comes from the energy gap of $H(\mathbf{k}, \mathbf{r})$.

[12]The kernel $\ker(\hat{h}^\dagger)$ is the subspace that is annihilated by the operator $\hat{h}$. In general, this is an infinite dimensional space. However, one can define the *Laplacian* $\Delta = \hat{h}^\dagger \hat{h}$, and it can be shown that the chirality is given by the finite difference

$$
\begin{aligned}
\mathrm{ind}(\hat{H}) &= \dim \ker(\Delta) - \dim \ker(\hat{h}) & (5.3.15)\\
&= \dim \ker(\Delta) - \dim \ker(\hat{h}) & (5.3.16)
\end{aligned}
$$

where the Laplacian operator Δ is *Fredholm* and has finite dimensional kernel [Nakahara, 1990; Berline, Getzler and Vergne, 2004]. This is guaranteed by the ellipticity of $\hat{H}$, i.e. the bulk gap of $H(\mathbf{k}, \mathbf{r})$.

representing a winding number in the stable homotopy group $\pi_{2d-1}(U)$. It is a consequence of the Atiyah-Singer index theorem [Atiyah and Singer, 1968; Nakahara, 1990; Berline, Getzler and Vergne, 2004] that the analytical and topological index equate [Jackiw and Rebbi, 1976; Jackiw and Rossi, 1981; Goldstone and Wilczek, 1981; Weinberg, 1981; Witten, 1982; Davis, Davis and Perkins, 1997; Volovik, 2003]

$$\text{ind}(\hat{H}) = \frac{(d-1)!}{(2d-1)!(2\pi i)^d} \int_{BZ \times \mathbb{S}^{d-1}} \text{Tr}[(q dq^\dagger)^{2d-1}] \tag{5.3.17}$$

where $q(\mathbf{k}, \mathbf{r})$ are the off-diagonal component of the flattened Hamiltonian (5.2.18).

Majorana zero mode It has been already shown previously in section 4.1 the appearance of low energy Majorana bound excitations at (i) the ends of the Kitaev's 1D superconducting chain (fig.4.1 and 4.2), (ii) a 1D SC-FM domain wall on top of a QSHI edge (fig.4.3), (iii) a π-flux plaquette of the Kitaev's honeycomb model (fig.4.4), and (iv) a vortex core of a 2D chiral p-wave superconductor (fig.4.5). In the subsequent section 4.2, low energy Majorana bound excitations are shown to be associated with hedgehog defects $\vec{n}(\mathbf{r})$ in eq.(4.2.8) and fig.4.9 (or mass term $m\Gamma(\mathbf{r})$ in (4.2.10)) and can be realized theoretically by a TI-SC heterostructure fig.4.8.

All these are zero energy excitations associated to topological point defects described by semi-classical BdG Hamiltonians in symmetry class D. The low energy Hamiltonian away from the point singularity can all be expressed phenomenologically as the Dirac-type model

$$H(\mathbf{k}, \mathbf{r}) = \hbar v_f \sum_{i=1}^{d} k_i \gamma_i + m\Gamma(\mathbf{r}) \tag{5.3.18}$$

where the mass term $m\Gamma(\mathbf{r})$ lives in the classifying space $\mathcal{R}_{2-d}$.[13] The existence of robust Majorana zero mode is indicated by the winding of the mass term $m\Gamma$: $\mathbb{S}^2 \to \mathcal{R}_{2-d}$ in the stable homotopy group $\pi_{d-1}(\mathcal{R}_{2-d}) = \mathbb{Z}_2$.

In general, all point defects of symmetry class D in any dimensions are characterized by the Chern-Simons $\mathbb{Z}_2$-invariant (5.2.22)

$$\nu \equiv \frac{2}{d!} \left(\frac{i}{2\pi} \right)^d \int_{BZ \times \mathbb{S}^{d-1}} \mathcal{Q}_{2d-1} \qquad (\text{mod } 2) \qquad (5.3.19)$$

In particular a $2\pi m$ pairing vortex across a chiral 2D TR-breaking superconductor with Chern invariant $p \in \mathbb{Z}$ is a point defect in 2D with Chern-Simons invariant

$$\nu \equiv pm \qquad (\text{mod } 2) \qquad (5.3.20)$$

And this counts the number of Majorana zero modes bounded at the vortex core.

Majorana Kramer's doublet   The excitation appears at a topological point defect in the time reversal symmetric class DIII. Kramer's theorem requires any Majorana zero modes to be doubly degenerate and have opposite chirality according to $\Pi = i\Theta\Xi$. This theoretically arise at a π-Josephson junction in proximity with the edge of an QSHI [Fu and Kane, 2008, 2009a]. Its topology is characterized by another Chern-Simons $\mathbb{Z}_2$-invariant (5.2.26)

$$\tilde{\nu} \equiv \frac{1}{d!} \left(\frac{i}{2\pi} \right)^d \int_{BZ \times \mathbb{S}^{d-1}} \mathcal{Q}_{2d-1} \qquad (\text{mod } 2) \qquad (5.3.21)$$

---

[13]For d=3, it has been worked out that $\mathcal{R}_{2-d}$ is the symmetric space U/O. (See below eq.(4.2.10))
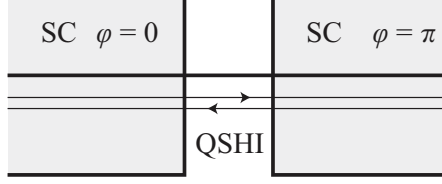
Figure 5.11: A π-Josephson junction on top of a quantum spin Hall edge. A Kramer's doublet of Majorana zero modes is bounded at the interface.

which differs from (5.3.19) by a factor of 2 and the Berry connection is chosen with a TR gauge constraint

$$w_{mn}(\mathbf{k}, \mathbf{r}) = \langle u_m(\mathbf{k}, \mathbf{r})|\Theta u_n(-\mathbf{k}, \mathbf{r})\rangle \equiv \text{constant} \tag{5.3.22}$$

where u_m are valence states of the band Hamiltonian. In one dimension such as the QSHI edge, this can be simplified into a Pfaffian invariant [Qi, Hughes and Zhang, 2008]

$$(-1)^{\tilde{\nu}} = \frac{\text{Pf}[q(\pi)]}{\text{Pf}[q(0)]} \sqrt{\frac{\det[q(0)]}{\det[q(\pi)]}} \tag{5.3.23}$$

where $q(k)$ is the off-diagonal component of the 1D BdG Hamiltonian $H(k)$ that anticommute with $\Pi = \tau_z$, and the sign of the square root is chosen so that the branch $\sqrt{\det[q(k)]}$ is continuously defined in between $k = 0, \pi$.

5.3.3   Adiabatic pumps

An adiabatic pump here refers to a time dependent cycle with multiple point defects. It has dimension $\delta = d - D = 0$. $(\mathbf{k}, \mathbf{r}, t)$ lives in the base space $T^d \times (\wedge^{N-1}\mathbb{S}^{d-1}) \times \mathbb{S}^1$. $\wedge^{N-1}\mathbb{S}^{d-1}$ is a bouquet of $(N-1)$ spheres, each enclosing a point defect in real d-space and $\mathbb{S}^1$ represents the time cycle. The topological signature of an adiabatic cycle involve pumping charge in between point defects or fermion parity in between

Symmetry	Topological classification	Adiabatic Pump
A	$\mathbb{Z}$	Charge
AI	$\mathbb{Z}$	Charge
BDI	$\mathbb{Z}_2$	Fermion Parity
D	$\mathbb{Z}_2$	Fermion Parity
AII	$2\mathbb{Z}$	Charge Kramers Doublet

Table 5.8: Symmetry classes that support non trivial charge or fermion parity pumping cycles.

pairs. The possible topological cycles are listed in table 5.8. The 1D Thouless pump [Thouless, 1983] and the 2D Laughlin's argument [Laughlin, 1981] of integer QHE are typical examples that falls into the catergory of charge pumps. Fermion parity pumps (a.k.a. braidless operations) were previously studied in detail in section 4.3. The topologies were characterized by a Pontryagin invariant $[\mathbb{S}^2 \times \mathbb{S}^1, \mathbb{S}^2]_{\mathrm{deg}=1} = \mathbb{Z}_2$ (or the stable homotopy $\pi_3(U/O)$) with the help of a Dirac-type low energy model. In the framework of topological defect, it is characterized by the more general Fu-Kane $\mathbb{Z}_2$-invariant (5.2.25), although it could be computationally difficult to satisfies the PH gauge constraint (5.2.30).

In the 1D case (see fig.4.12(B)), it is possible to apply (5.2.25) because the Dirac-type Hamiltonian of the chiral 1D superconductor

$$H_{BdG}(k, \varphi(t)) = (t \cos k - \mu)\tau_z + \Delta \sin k(\cos \varphi(t)\tau_x + \sin \varphi(t)\tau_y) \qquad (5.3.24)$$

is time independent at $k = 0, \pi$, which means that the PH gauge constraint (5.2.30) is automatically satisfied. Moreover, the Chern-Simons integral in the Fu-Kane formula is identically zero and therefore the adiabatic cycle is topologically charac-

terized by

$$\tilde{\nu} \equiv \frac{i}{2\pi} \int_0^\pi dk \int_0^{2\pi} d\varphi \, \text{Tr}(\mathcal{F}) \qquad (\text{mod } 2) \qquad (5.3.25)$$

which can be checked straightforwardly to be unity.

Chapter 6

Conclusion and outlook

Single particle topological phases of insulators and superconductors were studied,
the bulk-boundary correspondence was generalized and established in the context of
topological defects in band theory, and properties of gapless excitations associated
with interfaces and more general defects were investigated. Surface Dirac cone spec-
trum of the topological insulator $Bi_{1-x}Sb_x$ was derived using a tight binding model
in chapter 2. It exhibited the strong $\mathbb{Z}_2$ signature of the material as well as infor-
mation regarding inversion and mirror symmetry of the occupied states. Transport
characteristics of helical edge modes of QSHI were theoretically analyzed in a quan-
tum point contact in chapter 3. Critical conductivity behavior was predicted using
a spinful Luttinger liquid theory that incorporates electron-electron interactions.
Low energy Majorana quasiparticle excitations were shown to exist at certain point
defect configurations in a 3D superconductor in chapter 4. Interconnecting through
a non-local bulk parameter, they carried Ising non-Abelian statistics in $(3 + 1)D$.
The classification of insulators and superconductors was generalized to topological

defects in chapter 5. The topologies of defect band theory in different dimensions and symmetry classes were characterized by universal integral invariants. Physical situations in line and point defects as well as adiabatic cycles were proposed or revisited and the topological nature of their corresponding defect excitations were shown.

The study of topological phases in single particle theories was pioneered by Thouless, Kohmoto, Nightingale and den Nijs [1982] to understand the integer quantum Hall effect. The discovery of time reversal symmetric topological insulators [Kane and Mele, 2005a,b; Bernevig, Hughes and Zhang, 2006; Fu, Kane and Mele, 2007; Moore and Balents, 2007; Qi, Hughes and Zhang, 2008; Roy, 2006a,b] stimulated the theoretical developments in topological band theory [Schnyder *et.al.*, 2008, 2009; Kitaev, 2009; Teo and Kane, 2010b; Stone, Chiu and Roy, 2011]. The exploration of topological phases of matter is far from complete. Additionally there are numerous theoretical challenges in further understanding topological order and experimental challenges in realizing topological phases.

The fractional quantum Hall effect [Tsui, Stormer and Gossard, 1982] could not be understood without using a many-body theory that describes electron-electron interactions [Laughlin, 1983; Halperin, 1984]. This leads to a higher notion of topological order [Wen and Niu, 1990; Wen, 1995], such as a fractional quantum spin Hall insulator [Levin and Stern, 2009]. The theory of a *fractional* time reversal symmetric but spin-orbit coupled insulator has not been established. A non-perturbative framework incorporating time reversal symmetry (or other kind of symmetries) into an interacting many-body system is needed to describe a more general notion of

topological order.

Although the time reversal symmetric topological insulating phases in 2 and 3 dimensions are robust against interactions [Kane, 2008; Qi, Hughes and Zhang, 2008] due to charge or fermion parity conservation, the effect of interactions on single particle topological phases could be highly non-trivial in general in a 1D theory of Majorana fermions [Fidkowski and Kitaev, 2011; Turner, Pollmann and Berg, 2011], or be completely benign in an electromagnetic and gravitational response theory [Ryu, Moore and Ludwig, 2011]. The role of interactions for a bulk topological band theory for $d \geq 2$ or a defect theory in general needs to be carefully addressed and explored.

A lot of ARPES experiments have been carried out on the topological insulators Bi_2Se_3 and Bi_2Te_3 [Xia *et al.*, 2009a; Hor *et.al.*, 2009; Hsieh *et.al.*, 2009a,b]. Transport experiments are however more difficult due to the slightly conducting nature of the bulk material. 3D topological superconductors on the other hand are relatively uncharted experimentally. Cu-doped Bi_2Se_3 has been studied as a candidate [Hor *et.al.*, 2010; Fu and Berg, 2010; Wray *et.al.*, 2011]. It would be interesting to observe a Majorana Dirac surface excitation in an electronic system. The manipulation of Majorana bound states leads to Ising statistics. It would be interesting to carry out the proposals put forward by Fu and Kane [2008] and Alicea [2010] on superconductor heterostructures.

Bibliography

Aalseth, C.E., F.T. Avignone III, A. Barabash, F. Boehm, R.L. Brodzinski, J.I. Collar, P.J. Doe, H. Ejiri, S.R. Elliott, E. Fiorini, R.J. Gaitskell, G. Gratta, R. Hazama, K. Kazkaz, G.S. King III, R.T. Kouzes, H.S. Miley, M.K. Moe, A. Morales, J. Morales, A. Piepke, R.G.H. Robertson, W. Tornow, P. Vogel, R. A. Warner and J.F. Wilkerson, Mod. Phys. Lett. A **17**, 1475 (2002); H.V. Klapdor-Kleingrothaus, A. Dietz, H.L. Harney, I.V. Krivosheina, Mod. Phys. Lett. A **16**, 2409 (2001).

Ardonne, E., and K. Schoutens, Phys. Rev. Lett. **82**, 5096 (1999).

Akhmerov, A.R., J. Nilsson and C. W. J. Beenakker, Phys. Rev. Lett. **102**, 216404 (2009).

Alicea, J., Phys. Rev. B **81**, 125318 (2010).

Altland, A. and M. R. Zirnbauer, Phys. Rev. B **55**, 1142 (1997).

Alvarez-Gaumé, L., S. Della Pietra and G. Moore, Ann. of Phys. **163**, 288 (1985).

Arovas, D., J. R. Schrieffer, and F. Wilczek, Phys. Rev. Lett. **53**, 722 (1984).

Atiyah, M.F., *K Theory*, Westview Press (1994).

Atiyah, M.F. and I.M. Singer, Bull. Amer. Math. Soc. **69**, 322 (1963).

Atiyah, M.F. and I.M. Singer, Ann. Math. **87**, 484 (1968); Ann. Math. **87**, 546 (1968); Ann. Math. **93**, 119 (1971); Ann. Math. **93**, 139 (1971)

Atiyah, M.F. and G.B. Segal, Ann. Math. **87**, 531 (1968).

Atiyah, M.F., V.K. Patodi and I.M. Singer, Bull. Lond. Math. Soc. **5**, 229 (1973).

Avignone, F.T., S.R. Elliott, and J. Engel, Rev. Mod. Phys. **80**, 481 (2008).

Baxter, R.J., *Exactly solved models in statistical mechanics* Academic Press (1982).

Bellissard, J., A. van Elst and H. Schulz-Baldes, J. Math. Phys. **35**, 5373 (1994)

Berline, N., E. Getzler and M. Vergne, *Heat Kernels and Dirac Operators*, Springer, Berline (2004).

Bernevig, A., T. Hughes and S.C. Zhang, Science **314**, 1757 (2006).

Bernevig, A. and S.C. Zhang, Phys. Rev. Lett. **96**, 106802 (2006).

Berry, M.V., Proc. R. Soc. A **392**, 45 (1984).

Bertlmann, R.A., *Anomalies in Quantum Field Theory*, Oxford University Press, USA (2001)

Bloch, F., Z. Physik **52**, 555 (1929).

Bonderson, P., M. Freedman and C. Nayak, Phys. Rev. Lett. **101**, 010501 (2008).

Bott, R., Bull. Soc. Math. Fr. **84**, 251 (1956).

Caroli, C., P. G. de Gennes and J. Matricon, Phys. Lett. **9**, 307 (1964).

Cartan, E., Bull. Soc. Math. Fr. **54** 214 (1926); Bull. Soc. Math. Fr. **55**, 114 (1927).

Chaikin, P.M. and T.C. Lubensky, *Principles of Condensed Matter Physics*, Cambridge University Press, (2000).

Chern, S.S., Ann. Math. **47**, 85 (1946).

Chern, S.S. and J. Simons, Ann. Math. **99** 48 (1974).

Chung, S.B. and S.C. Zhang, Phys. Rev. Lett. **103**, 235301 (2009).

De Gennes, P.G., *Superconductivity of Metals and Alloys*, W.A. Benjamin, New York (1966).

Davis, S.C., A.C. Davis and W.B. Perkins, Phys. Lett. **408B**, 81 (1997).

Das Sarma, S., M. Freedman and C. Nayak, Phys. Rev. Lett. **94**, 166802 (2005).

Das Sarma, S., C. Nayak, S. Tewari, Phys. Rev. B **73**, 220502(R) (2006).

Doplicher, S., R. Haag and J.E. Roberts, Comm. Math. Phys. **23**, 199 (1971).

Doplicher, S., R. Haag and J.E. Roberts, Comm. Math. Phys. **35**, 49 (1971).

Fidkowski, L. and A. Kitaev, Phys. Rev. B **83**, 075103 (2011).

Franz, M., Physics **3**, 24 (2010).

Freedman, M., M. Larsen and Z. Wang, Commun. Math. Phys. **227**, 605 (2002).

Freedman, M., M.B. Hastings, C. Nayak, X.L. Qi, K. Walker and Z. Wang, Phys. Rev. B **83**, 115132 (2011).

Fu, L., Phys. Rev. Lett. **104**, 056402 (2010).

Fu, L. and E. Berg, Phys. Rev. Lett. **105**, 097001 (2010).

Fu, L. and C.L. Kane, Phys. Rev. B **74**, 195312 (2006).

Fu, L., C.L. Kane and E.J. Mele, Phys. Rev. Lett. **98**, 106803 (2007).

Fu, L. and C.L. Kane, Phys. Rev. B **76**, 045302 (2007).

Fu, L. and C.L. Kane, Phys. Rev. Lett. **100**, 096407 (2008).

Fu, L. and C.L. Kane, Phys. Rev. B **79**, R161408 (2009).

Fu, L. and C.L. Kane, Phys. Rev. Lett. **102**, 216403 (2009).

Fukui, T., T. Fujiwara and Y. Hatsugai, J. Phys. Soc. Jpn. **77**, 123705 (2008).

Fukui, T. and Y. Hatsugai, J. Phys. Soc. Jpn. **76**, 053702 (2007).

Furusaki, A. and N. Nagaosa, Phys. Rev. B **47**, 4631 (1993).

Greiter, M., X. G. Wen and F. Wilczek, Nucl. Phys. B **374**, 567 (1992).

Goldstone, J. and F. Wilczek, Phys. Rev. Lett. **47**, 986 (1981).

Golin, S., Phys. Rev. **166**, 643 (1968).

Golin, S., Phys. Rev. **176**, 830 (1968).

Haldane, F.D.M., Phys. Rev. Lett. **61**, 2015 (1988).

Halperin, B.I., Phys. Rev. B **25**, 2185 (1982).

Halperin, B.I., Phys. Rev. Lett. **52**, 1583 (1984).

Hasan M.Z., and C.L. Kane, Rev. Mod. Phys. **82**, 3045 (2010).

Hastings, M.B. and T.A. Loring, J. Math. Phys. **51**, 015214 (2010).

Hatcher, A., *Algebraic topology*, Cambridge University Press (2002).

Hirahara, T., K. Miyamoto, I. Matsuda, T. Kadono, A. Kimura, T. Nagao, G. Bihlmayer, E.V. Chulkov, S. Qiao, K. Shimada, H. Namatame, M. Taniguchi and S. Hasegawa, Phys. Rev. B **76**, 153305 (2007).

Hsieh, D., D. Qian, L. Wray, Y. Xia, Y.S. Hor, R.J. Cava and M.Z. Hasan, Nature **452**, 970 (2008).

Hsieh, D., Y. Xia, L. Wray, D. Qian, A. Pal, J.H. Dil, J. Osterwalder, F. Meier, G. Bihlmayer, C.L. Kane, Y. S. Hor, R. J. Cava and M. Z. Hasan, Science **323**, 919 (2009).

Hsieh, D., Y. Xia, D. Qian, L. Wray, J.H. Dil, F. Meier, J. Osterwalder, L. Patthey, J.G. Checkelsky, N.P. Ong, A.V. Fedorov, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R.J. Cava and M.Z. Hasan, Nature **460**, 1101 (2009).

Hsieh, D., Y. Xia, D. Qian, L. Wray, F. Meier, J.H. Dil, J. Osterwalder, L. Patthey, A.V. Fedorov, H. Lin, A. Bansil, D. Grauer, Y.S. Hor, R.J. Cava and M.Z. Hasan, Phys. Rev. Lett. **103**, 146401 (2009).

Hor, Y. S., A. Richardella, P. Roushan, Y. Xia, J. G. Checkelsky, A. Yazdani, M. Z. Hasan, N. P. Ong, and R. J. Cava, Phys. Rev. B **79**, 195208 (2009).

Hor, Y.S., A.J. Williams, J.G. Checkelsky, P. Roushan, J. Seo, Q. Xu, H.W. Zandbergen, A. Yazdani, N. P. Ong and R.J. Cava, Phys. Rev. Lett. **104**, 057001 (2010).

Hou, C.Y., E.A. Kim and C. Chamon, Phys. Rev. Lett. **102**, 076602 (2009).

Ivanov, D.A., Phys. Rev. Lett. **86**, 268 (2001).

Jackiw, R. and C. Rebbi, Phys. Rev. D **13**, 3398 (1976).

Jackiw, R. and P. Rossi, Nucl. Phys. B **190**, 681 (1981).

Jang, J., D.G. Ferguson, V. Vakaryuk, R. Budakian, S.B. Chung, P.M. Goldbart and Y. Maeno, Science **331**, 186 (2011).

Kane, C.L., Nature Physics **4**, 348 (2008).

Kane, C.L. and M.P.A. Fisher, Phys. Rev. B **46**, 15233 (1992).

Kane, C.L. and M.P.A. Fisher, Phys. Rev. Lett. **68**, 1220 (1992).

Kane, C.L. and E.J. Mele, Phys. Rev. Lett. **95**, 226801 (2005).

Kane, C.L. and E.J. Mele, Phys. Rev. Lett. **95**, 146802 (2005).

Kane, C.L. and J.E. Moore, Physics World **24**, 32 (2011).

Karoubi, M., *K-theory: an introduction*, Springer-Verlag (1978).

Kitaev, A., Phys. Usp. (suppl.) **44**, 131 (2001); arXiv:cond-mat/0010440 (2000).

Kitaev, A., Ann. Phys. **303**, 2 (2003).

Kitaev, A., Ann. Phys. **321**, 2 (2006).

Kitaev, A., AIP Conf. Proc. **1134**, 22 (2009).

Kohn, W. and L.J. Sham, Phys. Rev. **140**, A1133 (1965).

Koroteev, Y.M., G. Bihlmayer, E. V. Chulkov and S. Blügel, Phys. Rev. B **77**, 045428 (2008).

König, M., S. Wiedmann, C. Brüne, A. Roth, H. Buhmann, L. Molenkamp, X.L. Qi and S.C. Zhang, Science **318**, 766 (2007).

Kopnin, N.B. and M.M. Salomaa, Phys. Rev. B **44**, 9667 (1991).

Klitzing, K.v., G. Dorda and M. Pepper, Phys. Rev. Lett. **45**, 494 (1980).

Laughlin, R.B., Phys. Rev. B **23**, R5632 (1981).

R.B. Laughlin, 1983, Phys. Rev. Lett. **50**, 1395.

Lawson, H.B. and M.L. Michelsohn, *Spin Geometry*, Princeton, NJ (1989).

Lenoir, B., A. Dauscher, X. Devaux, R. Martin-Lopez, Y.I. Ravich, H. Scherrer and S Scherrer, in *Proceedings of Fifteenth International Conference on Thermoelectrics* IEEE, New York, p.1-13 (1996).

Levin, M. and A. Stern, Phys. Rev. Lett. **103**, 196803 (2009).

Liu, Y. and R.E. Allen, Phys. Rev. B **52**, 1566 (1995).

Majorana, E., Nuovo Cimento **5**, 171 (1937).

Milliken, F.P., C.P. Umbach, and R.A. Webb, Solid State Commun. **97**, 309 (1996).

Moon, K., H. Yi, C.L. Kane, S.M. Girvin and M.P.A. Fisher, Phys. Rev. Lett. **71**, 4381 (1993).

Moore G., and N. Read, Nucl. Phys. B **360**, 362 (1991).

Moore, J.E., Nature **464**, 194 (2010)

Moore, J.E. and L. Balents, Phys. Rev. B **75**, 121306(R) (2007).

Nakahara, M., *Geometry, Topology and Physics*, Adam Hilger, Bristol (1990).

Nayak, C. and F. Wilczek, Nucl. Phys. B **479**, 529 (1996).

Nayak, C., S.H. Simon, A. Stern, M. Freedman, and S. Das Sarma, Rev. Mod. Phys. **80**, 1083 (2008).

Nielssen, H. and N. Ninomiya, Phys. Lett. **130B**, 389 (1983).

Phillips, J.C. and L. Kleinman, Phys. Rev. **116**, 287 (1959).

Pontrjagin, L.S., Rec. Math. [Mat. Sbornik] N.S. **9**, 331 (1941).

Qi, X.L., T. Hughes and S.C. Zhang, arXiv:0802.3537 (2008).

Qi, X.L., T. Hughes, and S.C. Zhang, Phys. Rev. B **81**, 134508 (2010).

Qi, X.L., R. Li, J. Zang and S.C. Zhang, Science **323**, 1184 (2009).

Qi, X.L. and S.C. Zhang, Physics Today **63**, 33 (2010).

Qi, X.L. and S.C. Zhang, arXiv:cond-mat/1008.2026 (2010); accepted in Rev. Mod. Phys. (2011).

Ran, Y., Y. Zhang, and A. Vishwanath, Nat. Phys. **5**, 298 (2009).

Ran, Y., arXiv:cond-mat/1006.5454 (2010).

Read, N. and D. Green, Phys. Rev. B **61**, 10267 (2000).

Read, N., E. Rezayi, Phys. Rev. B **59** 8084 (1999).

Roy, R., Phys. Rev. B **79**, 195321 (2009); cond-mat/0604211 (2006).

Roy, R., Phys. Rev. B **79**, 195322 (2009); cond-mat/0607531 (2006).

Ryu, S., A. P. Schnyder, A. Furusaki and A.W.W Ludwig, New J. Phys. **12**, 065010 (2010).

Ryu, S., J.E. Moore, A.W.W. Ludwig, arXiv:cond-mat/1010.0936 (2010); accepted in Phs. Rev. B. (2011).

Sau, J.D., R.M. Lutchyn, S. Tewari and S. Das Sarma, Phys. Rev. Lett. **104**, 040502 (2010).

Schnyder, A.P., S. Ryu, A. Furusaki and A.W.W. Ludwig, Phys. Rev. B **78**, 195125 (2008).

Schnyder, A.P., S. Ryu, A. Furusaki and A.W.W. Ludwig, AIP Conf. Proc. **1134**, 10 (2009).

Service, R.F., Science **332**, 193 (2011).

Sheng, D.N., Z.Y. Weng, L. Sheng, and F.D.M. Haldane, Phys. Rev. Lett. **97**, 036808 (2006).

Slater, J.C. and G.F. Koster, Phys. Rev. **94**, 1498 (1954).

Steenrod, N., *The Topology of Fibre Bundles*, Princeton University Press (1999).

Stern, A.F., F. von Oppen and E. Mariani, Phys. Rev. B **70**, 205338 (2004).

Stern, A. and B.I. Halperin, Phys. Rev. Lett. **96**, 016802 (2006).

Stone, M. and S.B. Chung, Phys. Rev. B **73**, 014505 (2006).

Stone, M., C.K. Chiu and A. Roy, J. Phys. A: Math. Theor. **44**, 045001 (2011).

Ström, A. and H. Johannesson, Phys. Rev. Lett. **102**, 096806 (2009).

Su, W.P., J.R. Schrieffer and A.J. Heeger, Phys. Rev. B **22**, 2099 (1980).

Teo, J.C.Y., L. Fu and C.L. Kane, Phys. Rev. B **78**, 045426 (2008).

Teo, J.C.Y. and C.L. Kane, Phys. Rev. B **79**, 235321 (2009).

Teo, J.C.Y. and C.L. Kane, Phys. Rev. Lett. **104**, 046401 (2010).

Teo, J.C.Y. and C.L. Kane, Phys. Rev. B **82**, 115120 (2010).

Thouless, D.J., M. Kohmoto, M.P. Nightingale and M. den Nijs, 1982, Phys. Rev. Lett. **49**, 405.

Thouless, D.J., Phys. Rev. B **27**, 6083 (1983).

Tsui, D.C., H.L. Stormer and A. C. Gossard, Phys. Rev. Lett. **48**, 1559 (1982).

Turner, A.M., F. Pollmann and E. Berg, Phys. Rev. B **83**, 075102 (2011).

Volovik, G.E. and V.M. Yakovenko, J. Phys.: Condens. Matter **1**, 5263 (1989).

Volovik, G.E., JETP Lett. **70**, 609 (1999).

Volovik, G.E., *The Universe in a Helium Droplet*, Clarendon, Oxford (2003).

Volovik, G.E., JETP Lett. **90**, 587 (2009).

Wang, Z., X.L. Qi and S.C. Zhang, New J. Phys. **12**, 065007 (2010).

Weinberg, E.J., Phys. Rev. D **24**, 2669 (1981).

Wen, X.G., Advances in Physics **44**, 405.

Wen, X.G. and Q. Niu, Phys. Rev. B **41**, 9377 (1990).

Wilczek, F., Phys. Rev. Lett. **48**, 1144 (1982).

Wilczek, F. and A. Zee, Phys. Rev. Lett. **51**, 2250 (1983)

Wilczek, Nat. Phys. **5**, 614 (2009).

Witten, E., Phys. Lett. 117B, 324 (1982); J. Diff. Geom. 17, 611 (1982).

Wray, L.A., Y. Xia, S.Y. Xu, D. Qian, A.V. Fedorov, H. Lin, A. Bansil, Y.S. Hor, R.J. Cava, L. Fu and M.Z. Hasan, arXiv:cond-mat/1104.4325 (2011); accepted in Phys. Rev. B. (2011);

Wu, C., B.A. Bernevig, S.C. Zhang, Phys. Rev. Lett. **96**, 106401 (2006).

Xia, Y., D. Qian, D. Hsieh, L. Wray, A. Pal, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava and M. Z. Hasan, Nat. Phys. **5**, 398 (2009).

Xu, C. and J.E. Moore, Phys. Rev. B. **73**, 045332 (2006).

Yi, H. and C.L. Kane, Phys. Rev. B **57**, R5579 (1998); H. Yi, Phys. Rev. B **65**, 195101 (2002).

Zak, J., Phys. Rev. **134**, A1602 (1964)

Surface states and topological invariants in three-dimensional topological insulators: Application to Bi$_{1-x}$Sb$_x$

Jeffrey C. Y. Teo, Liang Fu, and C. L. Kane

Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

We study the electronic surface states of the semiconducting alloy bismuth antimony (Bi$_{1-x}$Sb$_x$). Using a phenomenological tight-binding model, we show that the Fermi surface for the 111 surface states encloses an odd number of time-reversal-invariant momenta (TRIM) in the surface Brillouin zone. This confirms that the alloy is a strong topological insulator in the (1;111) $\mathbb{Z}_2$ topological class. We go on to develop general arguments which show that spatial symmetries lead to additional topological structure of the bulk energy bands, and impose further constraints on the surface band structure. Inversion-symmetric band structures are characterized by eight $\mathbb{Z}_2$ "parity invariants," which include the four $\mathbb{Z}_2$ invariants defined by time-reversal symmetry. The extra invariants determine the "surface fermion parity," which specifies which surface TRIM are enclosed by an odd number of electron or hole pockets. We provide a simple proof of this result, which provides a direct link between the surface-state structure and the parity eigenvalues characterizing the bulk. Using this result, we make specific predictions for the surface-state structure for several faces of Bi$_{1-x}$Sb$_x$. We next show that mirror-invariant band structures are characterized by an integer "mirror Chern number" $n_\mathcal{M}$, which further constrains the surface states. We show that the sign of $n_\mathcal{M}$ in the topological insulator phase of Bi$_{1-x}$Sb$_x$ is related to a previously unexplored $\mathbb{Z}_2$ parameter in the L point $\mathbf{k} \cdot \mathbf{p}$ theory of pure bismuth, which we refer to as the "mirror chirality" η. The value of η predicted by the tight-binding model for bismuth disagrees with the value predicted by a more fundamental pseudopotential calculation. This explains a subtle disagreement between our tight-binding surface-state calculation and previous first-principles calculations of the surface states of bismuth. This suggests that the tight-binding parameters in the Liu-Allen model of bismuth need to be reconsidered. Implications for existing and future angle-resolved photoemission spectroscopy (ARPES) experiments and spin-polarized ARPES experiments will be discussed.

I. INTRODUCTION

A topological insulator is a material with a bulk electronic excitation gap generated by the spin-orbit interaction, which is topologically distinct from an ordinary insulator.[1–8] This distinction, characterized by a $\mathbb{Z}_2$ topological invariant, necessitates the existence of gapless electronic states on the sample boundary. In two dimensions, the topological insulator is a quantum spin Hall insulator,[1–3,8] which is a close cousin of the integer quantum Hall state. The edge states predicted for this phase have recently been observed in transport experiments on HgCdTe quantum wells.[9] In three dimensions there are four $\mathbb{Z}_2$ invariants characterizing a time-reversal-invariant band structure.[5–7] One of these distinguishes a strong topological insulator, which is robust in the presence of disorder. The strong topological insulator is predicted to have surface states whose Fermi surface encloses an odd number of Dirac points and is associated with a Berry's phase of π. This defines a topological metal surface phase, which is predicted to have novel electronic properties.[7,10,11]

In Ref. 12 we predicted that the semiconducting alloy Bi$_{1-x}$Sb$_x$ is a strong topological insulator using a general argument based on the inversion symmetry of bulk crystalline Bi and Sb. The surface states of Bi have been studied for several years. Experimentally there are several photoemission studies of Bi crystals and films which have probed the surface states.[13–22] There are fewer studies of Bi$_{1-x}$Sb$_x$,[23] but in a very recent work, Hsieh *et al.*[24] mapped the (111) sur-

face states and verified the topological structure predicted for a strong topological insulator.

First-principles calculations provide a clear picture of the surface-state structure of Bi,[19–22,25] which captures many of the experimental features, including their spin structure.[22] For the alloy Bi$_{1-x}$Sb$_x$, one expects the surface states to evolve smoothly from Bi, at least for small x. The alloy presents two difficulties for these calculations, though. First, since these calculations must be done on relatively thin slabs, features near the small band gap are inaccessible because finite-size quantization mixes the bulk and surface states. Moreover, describing the alloy would require some kind of mean-field treatment of the substitutional disorder.

In this paper we study the surface states of Bi$_{1-x}$Sb$_x$ first by developing a phenomenological tight-binding model which can be solved numerically and then by developing general arguments that exploit spatial symmetries and explain a number of model independent features of the surface states. Our phenomenological tight-binding model is based on an interpolation of a model developed by Liu and Allen.[26] This model has the advantage that it can be solved in a semi-infinite geometry, which allows the surface-state features near the small band gap to be calculated. Our aim is not to perform a quantitatively accurate calculation of the surface states, but rather to provide a concrete calculation in which robust, model independent features of the surface states can be identified and characterized. Here we list our main conclusions:

(1) We find that the Fermi surface of the 111 surface of $Bi_{1-x}Sb_x$ consists of an electron pocket centered around the $\bar{\Gamma}$ point and six elliptical hole pockets centered a point in between $\bar{\Gamma}$ and the $\bar{M}$ point. (Here the bar refers to symmetry points in the 111 surface Brillouin zone.) This is similar to the surface states in Bi. Unlike the alloy, however, Bi has bulk states at the Fermi energy: hole states near $\bar{\Gamma}$ and electron states near $\bar{M}$. This calculation verifies the topological structure of the surface states predicted in Ref. 12. In that work we showed that the four $\mathbb{Z}_2$ invariants $(\nu_0; \nu_1\nu_2\nu_3)$ characterizing the valence bands of pure Bi and Sb are $(0;000)$ and $(1;111)$, respectively. The semiconducting alloy $Bi_{1-x}Sb_x$ was argued to be in the same class as Sb, which is a strong topological insulator. These invariants determine the number of surface bands crossing the Fermi energy modulo 2 between each pair of time-reversal-invariant momenta (TRIM) in the surface Brillouin zone. Specifically, it predicts that for the 111 surface, an odd number of Fermi surface lines separate the $\bar{\Gamma}$ point from the three equivalent $\bar{M}$ points. This is consistent with both our calculation and with experiment.[24]

(2) We will show that for crystals with inversion symmetry, there is an additional topological structure in the bulk band structure, which further constrains the surface band structure. At each of the eight TRIM, Γ_i in the bulk Brillouin zone, the product of the parity eigenvalues of the occupied bands defines a *parity invariant* $\delta(\Gamma_i)$, which is a topological invariant in the space of inversion-symmetric Hamiltonians. The four $\mathbb{Z}_2$ invariants, which require only time-reversal symmetry are determined by these eight signs and determine the number of Fermi surface lines separating two surface TRIM. They do not, however, specify which of the TRIM are *inside* of the surface Fermi surface and which are *outside*. We will show that the bulk parity invariants $\delta(\Gamma_i)$ provide that information. Specifically, for each surface TRIM we will define the *surface fermion parity* as the parity of the number of Fermi lines that enclose that TRIM. This distinguishes the TRIM that are outside the Fermi surface from those that are inside a (single) electron or hole pocket. For a crystal terminated on an inversion plane, we will establish a theorem which relates the surface fermion parity to the bulk parity invariants. Thus, for inversion-symmetric crystals, the eight bulk parity invariants provide *more* information about the surface states than just the four $\mathbb{Z}_2$ invariants. We will give a simple proof of this theorem in the Appendix, which establishes a more direct connection between the bulk parity eigenvalues and the surface-state structure than that presented in Ref. 12. For the 111 surface of $Bi_{1-x}Sb_x$, our general theorem is consistent with both our surface-state calculation and with experiment. We will also apply this result to make predictions about the other surfaces of $Bi_{1-x}Sb_x$. In addition, our theorem has implications for inversion-symmetric crystals which are ordinary insulators. In particular, we will show that it has nontrivial implications for the surface states of pure Bi, whose valence band is in the trivial $(0;000)$ topological class.

(3) In addition to inversion symmetry, the crystal lattices of Bi and Sb have a mirror symmetry. We will show that the presence of mirror symmetry leads to a further topological classification of the bulk band structure in terms of an integer $n_{\mathcal{M}}$, which we refer to as a *mirror Chern number*. This integer is similar to the spin Chern number, which occurs in the quantum spin Hall effect when spin is conserved,[27] and its parity is related to the $\mathbb{Z}_2$ invariant.[28] The valence band of pure Bi, which has the $(0;000)$ $\mathbb{Z}_2$ class,[12] has $n_{\mathcal{M}}=0$. The semiconducting alloy is a topological insulator with $\mathbb{Z}_2$ class $(1;111)$. There are two possibilities for the mirror Chern number $n_{\mathcal{M}}=\pm 1$, however, which correspond to topologically distinct phases. We will show that the sign of $n_{\mathcal{M}}$ in the topological insulator phase further constrains the behavior of the surface states. The transition between the $(0;000)$ and $(1;111)$ classes in $Bi_{1-x}Sb_x$ occurs for small $x\sim.03$ because pure Bi is very close to a band inversion transition where the L_s valence band and L_a conduction band cross. The $\mathbf{k}\cdot\mathbf{p}$ theory of these states has been studied extensively in the literature[29–33] and has the form of a nearly massless *three-dimensional* Dirac point. We will show that the *change* $\Delta n_{\mathcal{M}}$ in the mirror Chern number at the band inversion transition is determined by a previously unexplored parameter in that theory: a sign $\eta=\pm 1$, which we will refer to as the *mirror chirality*. η is related to the sign of the g factor, which relates the magnetic moment to the angular momentum in a particular direction. For $\eta=+1$ the g factor is like that of a free electron, while for $\eta=-1$ it is anomalous. We will use this result to interpret our surface-state calculation and to provide guidance for how η can be measured. In addition to the Dirac point enclosed by the surface Fermi surface at $\bar{\Gamma}$, our tight-binding surface band calculations for both pure Bi and $Bi_{1-x}Sb_x$ predict that the six hole pockets also enclose Dirac points which reside at points along the line between $\bar{\Gamma}$ and $\bar{M}$. Unlike the Dirac points at the surface TRIM, the degeneracy at these Dirac points is not protected by time-reversal symmetry, but rather by mirror symmetry. This prediction is inconsistent with first-principles calculations of the surface states in Bi,[22,25] which do not find a band crossing inside the hole pocket. Since the Dirac point occurs above the Fermi energy, angle-resolved photoemission spectroscopy (ARPES) experiments do not directly probe this issue. Nonetheless, spin-resolved ARPES experiments on Bi provide evidence that the surface band structure is consistent with the first-principles calculations.[22] We will show that this inconsistency can be traced to the mirror chirality and the mirror Chern number. The mirror chirality in the topological insulator phase of $Bi_{1-x}Sb_x$ can be determined from the structure of the $\mathbf{k}\cdot\mathbf{p}$ perturbation theory of the energy bands in the vicinity of the L point in pure Bi. We find that the Liu-Allen model predicts that $n_{\mathcal{M}}=+1$. This value implies that the surface-state bands in the alloy cross in such a way as to establish the presence of the Dirac points in the hole pockets in agreement with our surface-state calculation. In contrast, we find that an earlier but more fundamental pseudopotential calculation by Golin[34] predicts that $n_{\mathcal{M}}=-1$. This value predicts that the bands do not cross and that there are no extra Dirac points, which is consistent with the presently available experimental results as well as first-principles calculations.[22,25] The Liu-Allen tight-binding parameters were chosen to reproduce the *energy* of the bands computed using first-principles calculations, incorporating available ex-
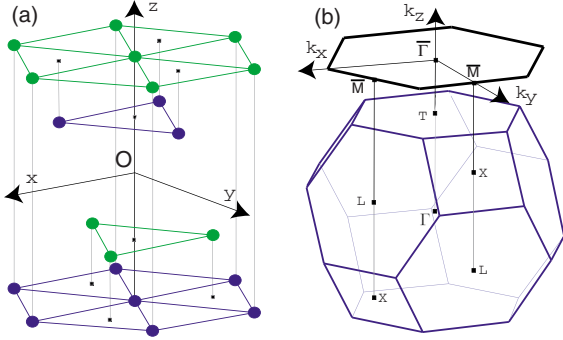
FIG. 1. (Color online) (a) Crystal structure of Bi. (b) Three-dimensional (3D) Brillouin zone and its projection onto the (111) surface. Also displayed is the choice of coordinate system throughout the paper: z is along the (111) direction, y is along the $\bar{\Gamma}$ to $\bar{M}$ direction, and O is a center of inversion.

perimental constraints. Therefore, there is no reason to expect that it gets $n_{\mathcal{M}}$ right. We conclude that the inconsistency in our surface-state calculation is an artifact of the Liu-Allen tight-binding model, which could be corrected with a suitable choice of new parameters.

The outline of the paper is as follows: In Sec. II we will review the salient features of bulk $Bi_{1-x}Sb_x$ and describe our phenomenological tight-binding model. In Sec. III we will describe our surface-state calculations for $Bi_{1-x}Sb_x$. In Sec. IV we will establish the relationship between the surface fermion parity and the bulk parity eigenvalues and use that result to analyze the surfaces of $Bi_{1-x}Sb_x$. In Sec. V we will discuss the mirror Chern number and show that it is related to the mirror chirality of the $\mathbf{k} \cdot \mathbf{p}$ theory of pure Bi. In Sec. VI we will conclude with a discussion of the relevance of our results to existing and future experiments. Finally, in the Appendix we provide a simple proof of the theorem relating the surface fermion parity to the bulk parity eigenvalues.

II. BULK $Bi_{1-x}Sb_x$

A. Introduction

Bismuth and antimony are group-V semimetals. They have the rhombohedral $A7$ structure shown in Fig. 1(a), which can be viewed as a distorted simple cubic lattice in which the triangular (111) lattice planes (which we will refer to as monolayers) are paired to form bilayers. The trivalent s^2p^3 atoms tend to form strong covalent bonds directed to the three nearest neighbors within a bilayer. Different bilayers are more weakly coupled. The primitive unit cell consists of two atoms in different monolayers, and each bilayer has a structure similar to a honeycomb lattice. The Brillouin zone for this lattice is shown in Fig. 1(b). It contains eight special points which are invariant under inversion and time reversal, denoted by Γ, T, and three equivalent L and X points.

Both Bi and Sb have a finite direct energy gap throughout the Brillouin zone, but they have a negative indirect gap. In Bi the conduction-band minimum at L is below the valence-band maximum at T, which gives rise to an anisotropic hole

pocket and three electron pockets with small effective masses.[31] At L the conduction-band minimum, which has even-parity L_s symmetry, nearly touches the valence-band maximum, with odd-parity L_a symmetry, forming a three-dimensional Dirac point with a small mass gap $E_g \approx 11$ meV. In Sb, the conduction-band minimum at L has L_a symmetry and is below the valence-band maximum at the lower-symmetry H point.

The alloy $Bi_{1-x}Sb_x$ retains the rhombohedral $A7$ crystal structure. The evolution of its band structure has been studied experimentally.[35,36] As x is increased from zero, two things happen. First, the small gap at L closes and then reopens. The L_s and L_a bands switch places, and the mass of the three-dimensional Dirac point changes sign. Second, the top of the valence band at T descends below the bottom of the conduction band, resulting in a semimetal-semiconductor transition. For $0.09 < x < 0.18$, the alloy is a direct-gap semiconductor, with a gap on order of 30 meV at the L points.

B. Topological invariants

Time-reversal-invariant band structures are classified topologically by four $\mathbb{Z}_2$ invariants.[5–7] In Ref. 12 we exploited inversion symmetry to show that these four invariants can be determined by the parity $\xi_m(\Gamma_i)$ of the occupied bands at the eight TRIM Γ_i, via the quantities

$$\delta(\Gamma_i) = \prod_n \xi_{2n}(\Gamma_i), \quad (2.1)$$

which we will refer to as parity invariants. Here the product includes each Kramer pair (which satisfy $\xi_{2n} = \xi_{2n-1}$) only once. For an inversion-symmetric crystal, all eight of the parity invariants are topological invariants in the following sense: If the crystal Hamiltonian is smoothly deformed, *preserving the inversion symmetry*, then the only way any of the $\delta(\Gamma_i)$ can change is if the gap at Γ_i goes to zero, so that states with opposite parities can be exchanged between the conduction and valence bands. If inversion symmetry is relaxed, then the eight invariants lose their meaning. However, in Ref. 12 we showed that provided *time-reversal symmetry* is preserved, four combinations of the $\delta(\Gamma_i)$ remain robust and define the four $\mathbb{Z}_2$ invariants denoted by $(\nu_0; \nu_1\nu_2\nu_3)$. The most important invariant, ν_0, distinguishes the strong topological insulator and survives even in the presence of disorder.[7,12] $(-1)^{\nu_0}$ is given simply by the product of all eight $\delta(\Gamma_i)$.

Pure Bi and Sb have inversion symmetry. The parity eigenvalues for inversion about the point O in Fig. 1(a) are tabulated in the literature.[26,34,37] Based on these data, we display $\delta(\Gamma_i)$ in Table I, along with the predicted $\mathbb{Z}_2$ invariants for pure Bi, pure antimony, and the alloy. The valence band of pure Bi is characterized by the trivial class (0;000), while antimony has the (1;111) class. The difference is due to the inversion of the L_s and L_a bands, which changes the sign of $\delta(L)$. The alloy inherits its topological class from antimony and is a strong topological insulator.

C. Pure Bi, Sb: Liu-Allen model

Liu and Allen[26] developed a third-neighbor tight-binding model for the electronic structure of Bi and Sb, which de-

TABLE I. Parity invariants $\delta(\Gamma_i)$ and $\mathbb{Z}_2$ topological invariants $(\nu_0; \nu_1 \nu_2 \nu_3)$ for bismuth, antimony, and $Bi_{1-x}Sb_x$ determined from the product of parity eigenvalues $\xi_m(\Gamma_i)$ at each bulk TRIM Γ_i.

	$\delta(\Gamma)$	$\delta(L)$	$\delta(T)$	$\delta(X)$	$(\nu_0; \nu_1 \nu_2 \nu_3)$
Bismuth	-1	-1	-1	-1	$(0;000)$
Antimony	-1	1	-1	-1	$(1;111)$
$Bi_{1-x}Sb_x$	-1	1	-1	-1	$(1;111)$

scribes the atomic s and p orbitals nearest to the Fermi energy. The Bloch Hamiltonian $\hat{H}(\mathbf{k}) = e^{-i\mathbf{k}\cdot\mathbf{r}} \mathcal{H} e^{i\mathbf{k}\cdot\mathbf{r}}$ has the form

$$\hat{H}(\mathbf{k}) = \begin{pmatrix} H_{11}(\mathbf{k}) & H_{12}(\mathbf{k}) \\ H_{21}(\mathbf{k}) & H_{22}(\mathbf{k}) \end{pmatrix}. \tag{2.2}$$

Here $H_{ab}(\mathbf{k})$ are 8×8 matrices describing the coupling between the $2s$ states and $6p$ states on the a and b sublattices of the crystal. The explicit form of these matrices is given in Tables IX and X in the appendix of Ref. 26.

$H_{11} = H_{22}$ describe the coupling within the same sublattice. These terms involve the on site energies E_s and E_p as well an on site spin-orbit coupling λ. The closest neighbor on the same sublattice is the third neighbor, which resides in the same monolayer as the origin. The third-neighbor hopping involves four parameters V_c'', with $c = ss$, $sp\sigma$, $pp\sigma$, and $pp\pi$, describing the hopping between the s and p states. Since further neighbor hopping is not included in this model, $H_{11}(\mathbf{k})$ and $H_{22}(\mathbf{k})$ describe decoupled monolayers and depend only on the momentum $\mathbf{q} = \mathbf{k}_\parallel$ in the plane of the monolayer.

$H_{12} = H_{21}^\dagger$ describes the coupling between the sublattices. These involve two terms: First-neighbor hopping terms V_c couples atoms within the same bilayer, and second-neighbor hopping terms V_c' couple atoms in neighboring bilayers. In the following it will be useful to separate these two contributions by writing $\mathbf{k} = (\mathbf{q}, k_z)$,

$$H_{12}(\mathbf{q}, k_z) = H_{12}^{(1)}(\mathbf{q}) e^{ik_z c_1} + H_{12}^{(2)}(\mathbf{q}) e^{-ik_z c_2}, \tag{2.3}$$

where c_1 and c_2 are the spacing between the monolayers within a bilayer and between different bilayers, and $\mathbf{q}$ and k_z are the momenta parallel and perpendicular to the surface. $H_{12}^{(1)}$ and $H_{12}^{(2)}$ can be extracted from Table X of Ref. 26 by noting that they are the terms which involve the parameters $g_0 - g_{12}$ and $g_{13} - g_{26}$, respectively.

The 12 hopping parameters and 3 on site parameters make a total of 15 parameters specifying this model. These were chosen to reproduce the energies predicted by first-principles calculations, as well as details of the band gaps and effective-mass tensors which are known experimentally. The values of the parameters for both Bi and Sb are listed in Table II of Ref. 26.

D. Tight-binding model for alloy

In order to describe the electronic structure of the alloy $Bi_{1-x}Sb_x$, we wish to develop a "virtual-crystal" approximation which treats the substitutional disorder in mean-field theory and results in a translationally invariant effective Hamiltonian. Since the regime of interest is $x \sim 0.1$, the ef-

fective Hamiltonian should be close to that of pure Bi. The effect of small x will be to modify the band energies, but not drastically change the wave functions. The effective Hamiltonian should reproduce two essential features: (1) the inversion of the L_s and L_a bands (which are nearly degenerate in pure Bi); and (2) the descent of the valence band at T below the conduction band at L, as x is increased, which leads to the transition between the semimetal and the semiconductor.

The simplest approach would be to simply interpolate between the tight-binding parameters for bismuth and antimony. For each of the 15 tight-binding parameters α_c, we could define

$$\alpha_c(x) = x\alpha_c^{Sb} + (1-x)\alpha_c^{Bi}. \tag{2.4}$$

However, for this simple interpolation the inversion between L_s and L_a occurs at a rather large value $x \sim 0.4$, which occurs after the semimetal-semiconductor transition. We found that this could be corrected if each of the hopping terms (but not the other terms) are revised such that

$$V_c(x) = xV_c^{Sb} + (1-x^2)V_c^{Bi}. \tag{2.5}$$

This approach is admittedly *ad hoc*, but it is sufficient for our purposes because it correctly accounts for the most important features of the band evolution. In Fig. 2 we plot the energies of T_{45}^+, L_s, and L_a as a function of x for this model. The qualitative behavior of the known band evolution is reproduced, including the decent of the hole pocket at T and the inversion of the conduction and valence bands at L. This should not, however, be interpreted as a quantitative description of the band evolution of $Bi_{1-x}Sb_x$.
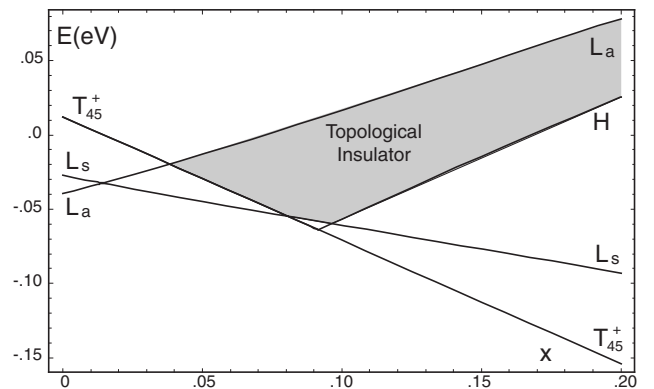
FIG. 2. Band evolution of interpolated tight-binding model using the parameters in Eqs. (2.4) and (2.5).

FIG. 3. (Color online) (a) Geometry for our surface-state calculations, which defines our co-ordinate system and specifies the spin directions of the $\bar{\Sigma}_1$ and $\bar{\Sigma}_2$ bands, which have mirror eigenvalues $+i$ and $-i$, respectively. (b) Brillouin zone for the (111) face of $Bi_{1-x}Sb_x$ with the electron pocket and six hole pockets predicted by our tight-binding calculation. (c) Surface band structure along the line between $\bar{\Gamma}$ and $\bar{M}$ predicted by the tight-binding model. The shaded regions are the bulk states projected onto the surface. (d) Schematic illustration of experimental surface band structure and Fermi surface probed by angle-resolved photoemission spectroscopy (Ref. 24). The top shows the Fermi surface in a slice of the Brillouin zone near $k_x = 0$, and the bottom shows the surface-state dispersion. Compared with (c), there are two additional bands near $\bar{M}$.

III. SURFACE STATES OF $Bi_{1-x}Sb_x$

In this section we describe our calculation of the 111 surface band structure for a semi-infinite lattice in the half plane $z < 0$ described by the interpolated tight-binding model described above. We begin with a brief discussion of our method, which is based on a transfer-matrix scheme,[38] and then go on to discuss the results.

A. Transfer-matrix method

The electronic states of a semi-infinite crystal can be represented as $\phi_{n,a}(\mathbf{q})$ in a basis of states which are plane waves with momentum $\mathbf{q}$ in the plane of the surface, but are localized on the $a = 1, 2$ monolayer of the nth bilayer. Each ϕ_{na} has eight components associated with the eight atomic orbitals. The time independent Schrödinger equation, written in this basis, may be expressed in the form

$$\begin{pmatrix} \phi_{n+1,1} \\ \phi_{n+1,2} \end{pmatrix} = T(\mathbf{q}, E) \begin{pmatrix} \phi_{n,1} \\ \phi_{n,2} \end{pmatrix}, \qquad (3.1)$$

where the transfer matrix is given by $T(\mathbf{q}, E) = t_{11}(\mathbf{q}, E) t_{22}(\mathbf{q}, E)$, with

$$t_{11} = \begin{pmatrix} H_{21}^{(2)-1}(E - H_{22}) & -H_{21}^{(2)-1} H_{21}^{(1)} \\ 1 & 0 \end{pmatrix}, \qquad (3.2)$$

and

$$t_{22} = \begin{pmatrix} H_{12}^{(1)-1}(E - H_{11}) & -H_{12}^{(1)-1} H_{12}^{(2)} \\ 1 & 0 \end{pmatrix}. \qquad (3.3)$$

Any bulk state is an eigenstate of the 16×16 transfer matrix with unimodular eigenvalues. For E within the energy gap, $T(\mathbf{q}, E)$ has exactly eight eigenvalues with modulus larger than 1. These correspond to states that decay exponentially in the $-z$ direction. $E(\mathbf{q})$ will correspond to a surface state localized at the top surface in Fig. 3(a) near $z = 0$ provided there is a linear combination of the decaying states which vanish on the monolayer $n = 0$, $a = 1$ just outside the surface: $\phi_{0,1} = 0$. The surface states are thus determined by forming an 8×8 matrix $M(\mathbf{q}, E)$ composed of the eight components of $\phi_{0,1}$ for each of the eight decaying states. $E(\mathbf{q})$ is then determined by solving $\det[M(\mathbf{q}, E)] = 0$.

B. Electronic structure of (111) surface

Figure 3(c) shows the energy spectrum of the (111) surface states of $Bi_{1-x}Sb_x$ for $x = 0.08$ calculated along the line connecting $\mathbf{q} = \bar{\Gamma} = 0$ to $\bar{M}$ along the $+\hat{y}$ axis using the transfer-matrix method for the interpolated tight-binding model. Figure 3(b) shows the Fermi surface. We find two bound surface states within the bulk energy gap. Along the line $q_x = 0$, these states are labeled by their symmetry under the mirror $\mathcal{M}(\hat{x})$, which takes x to $-x$. Since the mirror operation also operates on the spin degree of freedom, it is important to be specific about its definition. We write $M(\hat{x}) = PC_2(\hat{x})$, where P is inversion and $C_2(\hat{x})$ is a $180°$ counterclockwise rotation about the positive $\hat{x}$ axis. P does not affect the spin degree of freedom, but the C_2 rotation does. The resulting eigenvalues of $M(\hat{x})$ are $+i$ and $-i$, which we label as $\bar{\Sigma}_1$ and $\bar{\Sigma}_2$. These mirror eigenvalues are correlated with the spin S_x. For a free spin, eigenstates with $M(\hat{x}) = \pm i$ correspond to spin eigenstates with $S_x = \mp \hbar/2$. The surface states are not spin eigenstates, but on the line $k_x = 0, 0 < k_y < k_y(\bar{M})$, the expectation value of the spin satisfies $\langle \vec{S} \rangle \propto i \langle M(\hat{x}) \rangle \hat{x} \propto -(+)\hat{x}$, for $\Sigma_{1(2)}$, as indicated in Fig. 3(a).

The Fermi surface shown in Fig. 3(b) consists of electron and hole pockets. A single electron pocket surrounds $\bar{\Gamma}$. This

Fermi surface is nondegenerate, and opposite sides of the Fermi surface are Kramers pairs with opposite spins. The electronic states pick up a Berry's phase of π when they are adiabatically transported around the Fermi surface. This can be understood to be a consequence of the 360° rotation of the spin going around the Fermi surface. The Fermi surface is thus spin filtered, in the sense that the spin of the electron is correlated with its propagation direction, roughly satisfying $\langle \vec{S} \rangle \propto \hat{q} \times \hat{z}$ for an electron propagating in the $\hat{q}$ direction in the plane. In addition, there are six elliptical hole pockets centered along the six lines connecting $\bar{\Gamma}$ to $\bar{M}$. These are also nondegenerate, though unlike the electron pocket, the time reverse of a hole pocket is a different hole pocket. The crossing of the $\bar{\Sigma}_1$ and $\bar{\Sigma}_2$ bands is protected by the mirror symmetry for $q_x=0$. The degeneracy will be lifted for finite q_x, so the crossing describes a two-dimensional *Dirac* point, which is enclosed by the hole pocket.

C. Comparison with topological predictions

A single band of surface states connects the valence and conduction bands between $\bar{\Gamma}$ and $\bar{M}$ in Fig. 3(d). This confirms the topological predictions for the connectivity of the surface-state bands. In Ref. 7 we showed that the number of times $\Delta N(\Lambda_a, \Lambda_b)$ the surface states intersect the Fermi energy between two surface TRIM Λ_a and Λ_b satisfies

$$(-1)^{\Delta N(\Lambda_a, \Lambda_b)} = \pi(\Lambda_a)\pi(\Lambda_b), \tag{3.4}$$

where

$$\pi(\Lambda_a) = (-1)^{n_b}\delta(\Gamma_{a1})\delta(\Gamma_{a2}). \tag{3.5}$$

Here Γ_{a1} and Γ_{a2} are the two bulk TRIM which project to the surface TRIM Λ_a. The eight parity invariants $\delta(\Gamma_i)$, defined in Eq. (2.1), are products of parity eigenvalues. This definition of $\pi(\Lambda_a)$ differs slightly from the one introduced in Refs. 7 and 12 because of the additional factor $(-1)^{n_b}$. n_b is the number of occupied Kramers degenerate pairs of energy bands, which is equal to the number of terms in the product of Eq. (2.1). For $\mathrm{Bi}_{1-x}\mathrm{Sb}_x$, $n_b=5$. This factor does not affect $\Delta N(\Lambda_a, \Lambda_b)$ in Eq. (3.4). However, this modification simplifies our further results, discussed below.

For $\pi(\Lambda_a)\pi(\Lambda_b)=-1$, there will be an odd number of crossings between Λ_a and Λ_b, guaranteeing the presence of the gapless surface states. In the Appendix we will provide a derivation of this connection between the surface states and the bulk parity eigenvalues which is simpler and more direct than our previous proof.[12] This will show that with inversion symmetry the eight parity invariants $\delta(\Gamma_i)$ contain more information about the surface-state structure than just the number of crossings, a fact we will exploit in Sec. IV to make general predictions about the locations of electron and hole pockets in the surface Brillouin zone.

From Fig. 1, Table I, and Eq. (3.5), it can be seen that for the alloy,

$$\pi(\bar{\Gamma}) = -\delta(\Gamma)\delta(T) = -1, \tag{3.6}$$

$$\pi(\bar{M}) = -\delta(X)\delta(L) = +1. \tag{3.7}$$

This predicts that there should be an odd number of crossings between $\bar{\Gamma}$ and $\bar{M}$, which is confirmed both by our explicit calculation and, as we will discuss below, by experiment.

D. Comparison with experiment

Before comparing our calculation to experiment and other calculations, it is worthwhile to discuss what our calculation does *not* include. In addition to our approximate treatment of the alloy's bulk electronic structure, we have made no attempt to self-consistently describe the potential near the surface. This will be modified by relaxation of the bonds near the surface. More importantly, the population of the surface states determines the electric charge distribution near the surface, which leads to Hartree and exchange contributions to the potential. We assume that the surface is electrically neutral. We will argue in Sec. III that this means that the area of the electron pocket is equal to the total area of the six hole pockets. However, the potential due to a surface dipole layer is not included in our calculation. The effect of such a surface potential will be to modify the energies of the bands and perhaps to split off additional surface-state bands from the continuum. However, the topological connectivity of the surface-state bands will not be altered.

In their recent ARPES experiment, Hsieh *et al.*[24] measured the spectrum of $\mathrm{Bi}_{.9}\mathrm{Sb}_{.1}$ (111) surface states below E_F between $\bar{\Gamma}$ and $\bar{M}$. The observed spectrum, which we have sketched schematically in Fig. 3(d), resembles Fig. 3(c), though there are some important differences. As in Fig. 3(c), two surface-state bands emerge from the bulk valence band near $\bar{\Gamma}$. The first intersects the Fermi energy forming the electron pocket centered on $\bar{\Gamma}$, while the second intersects the Fermi energy forming a hole pocket. A third band crosses E_F from above, forming the opposite side of the hole pocket, and merges with the bulk valence band near $\bar{M}$. Unlike our calculation, the observed spectrum includes an additional electron pocket near $\bar{M}$. A Kramers degenerate pair of surface states is found in the gap at $\bar{M}$. Away from $\bar{M}$ these states split to form two surface bands, which both cross E_F near the end of the hole pocket. Thus there are a total of five bands crossing E_F between $\bar{\Gamma}$ and $\bar{M}$, which is consistent with the prediction for a (1;111) topological insulator. The discrepancy between our calculation and the experiment is most likely a consequence of our neglect of the self-consistent surface potential, which could lead to a Kramers pair of bound states to be split off from the conduction band at $\bar{M}$.

It is also instructive to compare our calculation with previous experimental and theoretical results for pure Bi. In Fig. 4(a) we show the surface-state spectrum for pure Bi calculated using the transfer-matrix method for the Liu-Allen tight-binding model. The number of band crossings is consistent with the trivial (0;000) topological structure of the Bi valence band. Since the Fermi energy of semimetallic Bi is
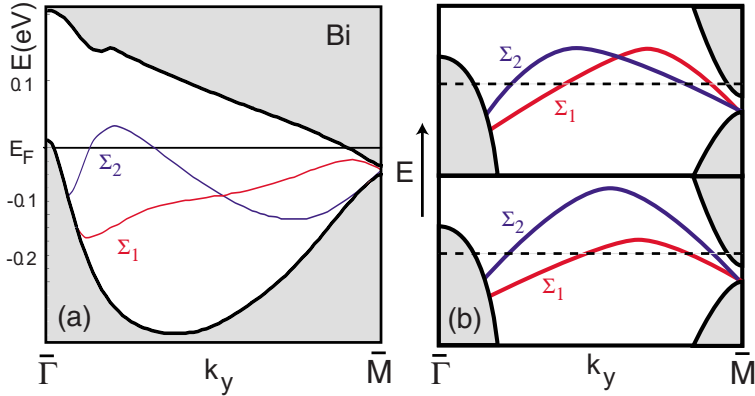
FIG. 4. (Color online) (a) Bi surface states between $\bar{\Gamma}$ and $\bar{M}$ calculated using tight-binding model. (b) Schematic picture of Bi bands in (a) in which Hartree effects raise the bands to accommodate charge neutrality. The crossing of Σ_1 and Σ_2 results in a Dirac point enclosed by a hole pocket. (c) Schematic picture without the crossing between Σ_1 and Σ_2, which resembles a first-principles calculation of surface states in Bi (Refs. 22 and 25).

fixed by the bulk, our calculated surface states violate surface charge neutrality: too many surface states are occupied, so the surface will have a negative charge. Hartree effects will push the surface states up in energy, but they will not alter the topological connectivity of the surface states. This allows us to deduce qualitative conclusions from the calculation.

First, as in our alloy calculation, two surface bands emerge from the bulk valence band near $\bar{\Gamma}$. These are also seen in photoemission experiments as well as first-principles calculations on pure Bi.[25] Moreover, the spin $\langle S_x \rangle$ of those surface states has been both calculated and measured using spin polarized ARPES.[22] We have checked that the spin direction predicted by our tight-binding calculation for each of these bands agrees with the experimental and first-principles theory results. Thus, the behavior near $\bar{\Gamma}$, including the ordering in which the $\bar{\Sigma}_2$ emerges first and forms the electron pocket, appears to be robust, with all calculations in agreement with each other and with experiment.

There is a discrepancy, however, between the tight-binding calculation and the first-principles calculation.[22,25] The crossing between the $\bar{\Sigma}_1$ and $\bar{\Sigma}_2$ bands in Figs. 3(c) and 4(a) is not found in the first-principles calculation. Since it is likely that this crossing would be pushed above the Fermi energy by Hartree corrections (so that the crossing occurs inside a hole pocket), the tight-binding model predicts that the hole pockets of Bi (111) enclose a Dirac point, as shown schematically in Fig. 4(b). The existence of this band crossing is not directly probed by ARPES, which only probes occupied states, though it could be probed using inverse photoemission. There is, however, *indirect* experimental evidence that the crossing does *not* occur. Spin polarized ARPES measurements[22] have measured the spin on both sides of the hole pocket. Though the signal appears weak, the sign of the spin is resolved and determined to be the same on both sides, indicating that there is no crossing, as shown schematically in Fig. 4(c). This agrees with the predictions of the first-principles calculations that both sides are in the same $\bar{\Sigma}_1$ band. In contrast, our tight-binding model predicts that the opposite sides of the hole pockets correspond to the $\bar{\Sigma}_1$ and $\bar{\Sigma}_2$ bands, which have opposite spins.

It thus appears likely that the prediction of the level crossing which implies that the hole pockets enclose a Dirac point is an artifact of the tight-binding model. This brings into question the related prediction of the tight-binding model

that the hole pockets of the alloy also enclose a Dirac point. In Sec. V we will argue that this artifact is a consequence of a subtle error in the Liu-Allen tight-binding model.

IV. INVERSION SYMMETRY AND THE SURFACE FERMION PARITY

An inversion-symmetric crystal can have no bulk electric polarization. In this section we show that this fact in combination with surface charge neutrality has nontrivial implications for the surface-state structure because it allows the *outside* of the surface Fermi surface to be unambiguously defined. It is then possible to define electron pockets to be regions in the surface Brillouin zone where an extra band is occupied and hole pockets as regions where an otherwise occupied band is empty. Charge neutrality dictates that the area of the electron pockets should equal that of the hole pockets. We will show that the locations of the electron and hole pockets in the surface Brillouin zone are topologically constrained by the bulk parity invariants $\delta(\Gamma_i)$. In addition to fixing the number of Fermi energy crossings, we find that $\delta(\Gamma_i)$ determine which TRIM are on the *inside* of an electron or hole pocket and which TRIM are on the *outside*. We define the surface fermion parity, which specifies whether a given surface TRIM is enclosed by an even or an odd number of Fermi lines. We will begin with a general discussion of the relationship between the surface fermion parity to the bulk parity invariants. We will then apply our general result to the surfaces of $Bi_{1-x}Sb_x$ and Bi.

A. Surface fermion parity

The total surface charge density may be expressed as a sum over the surface Brillouin zone (SBZ),

$$\sigma = e \int_{SBZ} \frac{d^2q}{(2\pi)^2} N(\mathbf{q}), \qquad (4.1)$$

where the surface fermion number $N(\mathbf{q})$ represents the excess charge in the vicinity of the surface due to states with momentum $\mathbf{q}$ in the plane of the surface. If we assume that the bulk Fermi energy is inside the gap, then there will be two contributions, $N(\mathbf{q}) = N_{bulk}(\mathbf{q}) + N_{surface}(\mathbf{q})$. $N_{surface}(\mathbf{q})$ is an integer which counts the occupied discrete surface states inside the energy gap. $N_{bulk}(\mathbf{q})$ is the total surface charge in the continuum valence-band states. For a crystal with inver-

FIG. 5. Two inequivalent inversion centers c and c' in an inversion-symmetric crystal, which differ by half a lattice vector. The parity eigenvalues of Bloch state at momentum $k = \pi/R$ with inversion center chosen at c and c' are different. Crystals terminated at c and c' will have surface charges that differ by an odd integer.

sion symmetry, there can be no bulk electric polarization, and $N_{\text{bulk}}(\mathbf{q})$ will also be quantized. In the Appendix we will show that it must be an integer.[39]

The integer values of $N(\mathbf{q})$ allow us to unambiguously define the "outside" of the surface Fermi surface to be the region for which $N(\mathbf{q}) = 0$. $N(\mathbf{q}) = +(-)1$ define electron (hole) pockets. $N(\mathbf{q}) = +(-)2$ is a double electron (hole) pocket, and so on. From Eq. (4.1), charge neutrality implies that the total area of the electron pockets equals that of the hole pocket provided that the double pockets are appropriately counted.

Kramers' theorem requires that the surface states be two-fold degenerate at the TRIM $\mathbf{q} = \Lambda_a$ in the surface Brillouin zone. Provided the Fermi energy is not exactly at the degeneracy point, this means that $N_{\text{surface}}(\Lambda_a)$ is even, so that the *parity* of $N(\Lambda_a)$ is equal to the parity of $N_{\text{bulk}}(\Lambda_a)$. In the Appendix we will show that the surface fermion parity is determined by the bulk parity invariants,

$$(-1)^{N(\Lambda_a)} \equiv \pi(\Lambda_a) = (-1)^{n_b} \delta(\Gamma_{a1}) \delta(\Gamma_{a2}). \qquad (4.2)$$

Equation (4.2) determines whether the TRIM Λ_a is enclosed by a single (or odd number) of Fermi lines, or whether it is outside the Fermi surface (or enclosed by an even number). In the special case that the Fermi energy is exactly at a Dirac point at Λ_a, Λ_a should be interpreted to be inside an electron (or hole) pocket with vanishing size.

Equation (4.2) is a central result of this paper which provides information about the structure of the surface Fermi surface beyond that determined by the $\mathbb{Z}_2$ invariants $(\nu_0; \nu_1 \nu_2 \nu_3)$. We will show below that this result can have nontrivial consequences even in materials which are *not* topological insulators. For example, we will see that Eq. (4.2) constrains the surface states of pure Bi.

In order to apply Eq. (4.2), it is essential to use the parity eigenvalues associated with an inversion center in the plane on which the crystal is terminated. As a simple example, Fig. 5 shows a one-dimensional inversion-symmetric lattice, which has two distinct inversion points. In general, a three-dimensional inversion-symmetric crystal has eight distinct inversion centers, which are related to each other by *half* a Bravais lattice vector: $\mathbf{c}' = \mathbf{c} + \mathbf{R}/2$. The parity eigenvalues associated with inversion center $\mathbf{c}'$ will be related to those associated with $\mathbf{c}$ by

$$\xi'_m(\Gamma_i) = \xi_m(\Gamma_i) e^{i\Gamma_i \cdot \mathbf{R}} = \pm \xi_m(\Gamma_i). \qquad (4.3)$$

An inversion plane will contain four of those points. For a given surface orientation, there are two distinct parallel inversion planes. For a surface terminated on one of those

inversion planes, $\pi(\Lambda_a)$ does not depend on which of the four inversion centers within the inversion plane are used. This can be seen by noting that

$$\pi'(\Lambda_a) = \pi(\Lambda_a) \exp[i n_b (\Gamma_{a1} - \Gamma_{a2}) \cdot \mathbf{R}], \qquad (4.4)$$

where n_b is the number of occupied bands. When $\mathbf{c}$ and $\mathbf{c}'$ are in the plane of the surface, the dot product in the exponent is zero. Crystals terminated on inequivalent inversion planes, however, will have different $N(\Lambda_a)$. For odd n_b, $\pi'(\Lambda_a) = -\pi(\Lambda_a)$, so that the parity of $N(\Lambda_a)$ changes at all four Λ_a. Thus, changing the inversion plane amounts to filling (or emptying) a single surface band throughout the surface Brillouin zone. Since $N(\Lambda_a)$ depends on how the crystal is terminated, it is not a bulk property. However, $\Delta N(\Lambda_a, \Lambda_b) = N(\Lambda_a) - N(\Lambda_b) \bmod 2$ is a bulk property, which is determined by the $\mathbb{Z}_2$ invariants $(\nu_0; \nu_1 \nu_2 \nu_3)$.

B. Application to $Bi_{1-x}Sb_x$

We now apply our general result to $Bi_{1-x}Sb_x$ surfaces. In order to apply Eq. (4.2), it is necessary to identify the appropriate inversion centers. The eight inversion centers of the rhombohedral $A7$ lattice are the following:

(1) $\mathbf{c}_0 = 0$, the origin in Fig. 1, which is between two bilayers.

(2)–(4) $\mathbf{c}_{j=1,2,3} = \mathbf{a}_j/2$. Here $\mathbf{a}_j$ are the three rhombohedral primitive Bravais lattice vectors, which connect an atom to the nearest three atoms on the same sublattice of the neighboring bilayer.[13] These points are at the center of a nearest-neighbor bond in the middle of a bilayer.

(5)–(7) $\mathbf{c}_{ij} \equiv (\mathbf{a}_i + \mathbf{a}_j)/2$ for $i \neq j$. These three points are at the center of a second-neighbor bond between two bilayers.

(8) $\mathbf{c}_{123} = (\mathbf{a}_1 + \mathbf{a}_2 + \mathbf{a}_3)/2$, which is directly above the origin in Fig. 1, in the middle of a bilayer. For a given surface orientation, these inversion centers are divided into two groups of four, which reside in two possible cleavage planes.

In Ref. 13, the (111), (110), and (100) faces of Bi are discussed, where the Miller indices (mno) refer to the rhombohedral reciprocal-lattice vector $m\mathbf{b}_1 + n\mathbf{b}_2 + o\mathbf{b}_3$ with $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi \delta_{ij}$. In these cases the preferred cleavage plane is the one which minimizes the number of broken first-neighbor bonds. In Table II we list the four inversion centers in the cleavage plane for each of these faces. For comparison, we have also included the (111)′ face, which is terminated in the middle of a bilayer (breaking three nearest-neighbor bonds). Table II also shows how the bulk TRIM project onto the surface TRIM, using the notation $\Lambda_a = (\Gamma_{a1} \Gamma_{a2})$. These data, combined with Table II, are sufficient to determine the surface fermion parity $\pi(\Lambda_a)$ for both the alloy $Bi_{1-x}Sb_x$ (BiSb) and pure Bi for each surface, as shown in Table II.

First, consider the 111 surface. The parity eigenvalues quoted in the literature, which determined Eq. (2.1) in Table I, are with respect to an inversion center between two bilayers [point O in Fig. 1(a)]. Thus, for a crystal cleaved between two bilayers, $N(\Lambda_a)$ can be deduced by combining Eq. (4.2) with

TABLE II. For each crystal face (hkl), we list the four inversion centers $\mathbf{c_j}$ on the cleavage plane along with the projections relating the four surface TRIM Λ_a to the bulk TRIM $\Gamma_{a1,2}$. For each Λ we list the surface fermion parity $\pi(\Lambda_a)$ for both $\mathrm{Bi}_{1-x}\mathrm{Sb}_x$ and Bi. $\pi(\Lambda_a)$ is a product of parity invariants at $\Gamma_{a1,2}$.

Face	$\mathbf{c}_j$	$\Lambda_a = (\Gamma_{a1}\Gamma_{a2})$	$\pi_{\mathrm{BiSb}}(\Lambda_a)$	$\pi_{\mathrm{Bi}}(\Lambda_a)$
(111)	$\mathbf{c}_0\ \mathbf{c}_{12}$	$\bar{\Gamma} = (\Gamma T)$	-1	-1
	$\mathbf{c}_{13}\ \mathbf{c}_{23}$	$3\bar{M} = (LX)$	$+1$	-1
(111)′	$\mathbf{c}_1\ \mathbf{c}_2$	$\bar{\Gamma} = (\Gamma T)$	$+1$	$+1$
	$\mathbf{c}_3\ \mathbf{c}_{123}$	$3\bar{M} = (LX)$	-1	$+1$
(110)	$\mathbf{c}_0\ \mathbf{c}_3$	$\bar{\Gamma} = (\Gamma X)$	-1	-1
	$\mathbf{c}_{12}\ \mathbf{c}_{123}$	$\bar{X}_1 = (LL)$	-1	-1
		$\bar{X}_2 = (LT)$	$+1$	-1
		$\bar{M} = (XX)$	-1	-1
(100)	$\mathbf{c}_1\ \mathbf{c}_{13}$	$\bar{\Gamma} = (\Gamma L)$	-1	$+1$
	$\mathbf{c}_{23}\ \mathbf{c}_{123}$	$\bar{M} = (TX)$	$+1$	$+1$
		$2\bar{M}' = (LX)$	-1	$+1$

$$\pi(\Lambda_a) = -\,\delta(\Gamma_{a1})\delta(\Gamma_{a2}), \qquad (4.5)$$

as shown in Table II. This implies the surface Fermi surface encloses $\bar{\Gamma}$, but not $\bar{M}$, as shown schematically in Fig. 6(a). Equation (4.2) says nothing about either the hole pockets seen in experiment and our calculation or the double electron pocket at $\bar{M}$ observed in experiment[24] on $\mathrm{Bi}_{1-x}\mathrm{Sb}_x$ but not in our calculation. In order for the surface to be neutral, however, the Fermi energy must either be at a Dirac point at $\bar{\Gamma}$ (so that the Fermi surface has vanishing area) or there must also be compensating electron/hole pockets elsewhere in the

surface Brillouin zone (but not enclosing $\bar{M}$).

It is also instructive to first consider a (111)′ face cleaved *between* the monolayers in a bilayer, despite the fact that such a surface would likely be unstable. Since the origin $\mathbf{c}_0$ is not in the cleavage plane, the parity eigenvalues in Eq. (2.1) need to be modified using Eq. (4.4). This has the effect of changing the sign of all of the $\pi(\Lambda_a)$, so that

$$\pi'(\Lambda_a) = +\,\delta(\Gamma_{a1})\delta(\Gamma_{a2}). \qquad (4.6)$$

From Table II we thus conclude that the three $\bar{M}$ points are enclosed by the Fermi surface, but not $\bar{\Gamma}$, as shown in Fig. 6(b).

For the 110 surface the cleavage plane with one broken bond includes the origin $\mathbf{c}_0$. Thus $\pi(\Lambda_a)$ can be determined with Eq. (4.5) along with the projections of the bulk TRIM shown in Table II. This leads to the predictions for the surface Fermi surface shown in Fig. 6(c). Experimental data for this face of $\mathrm{Bi}_{1-x}\mathrm{Sb}_x$ are currently unavailable. However, it is instructive to compare this prediction with experiments on pure Bi. In Ref. 15, single hole pockets are clearly seen at $\bar{\Gamma}$ and $\bar{M}$, and at X_1 single surface Dirac point is present inside the bulk gap. The situation at $\bar{X}_2$ is obscured due to the overlap of the bulk conduction and valence bands at L and T.

For the 100 surface the cleavage plane with one broken bond does not include $\mathbf{c}_0$. Thus, as was the case for the (111)′ surface, the surface fermion parity follows from Eq. (4.6). The surface Brillouin zone shown in Fig. 6(d) has TRIM $\bar{\Gamma}$, $\bar{M}$, and two equivalent $\bar{M}'$. Again, there are presently no data for this surface of $\mathrm{Bi}_{1-x}\mathrm{Sb}_x$. The (100) face of pure Bi is discussed in Ref. 20 and appears to be consistent with the prediction of Table II that none of the TRIM are enclosed by a Fermi surface.
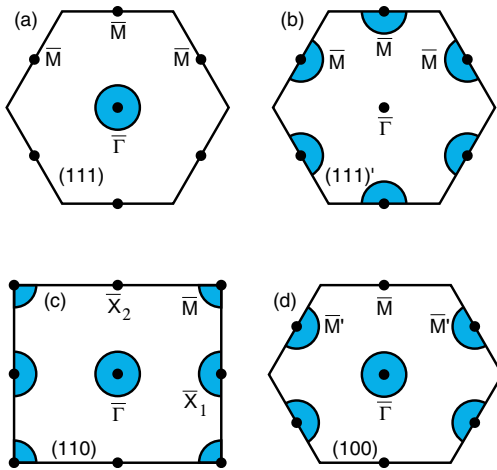
FIG. 6. (Color online) Schematic diagram showing which surface TRIM are enclosed by an odd number of electron or hole pockets for different faces of $\mathrm{Bi}_{1-x}\mathrm{Sb}_x$ predicted by the surface fermion parity in Table II. (a)–(d) show the (111), (111)′, (110), and (100) faces. The (111)′ surface is a hypothetical surface cleaved in the middle of a bilayer.

V. MIRROR CHERN NUMBER AND THE MIRROR CHIRALITY AT THE *L* POINT OF BISMUTH

In this section we will explore the consequences of mirror symmetry on the band structure of Bi and $Bi_{1-x}Sb_x$. This will address the disagreement between our calculation of the surface band structure and previous experimental and theoretical results. As discussed in Sec. III, the tight-binding model predicts that the hole pockets enclose Dirac points, while experiment and first-principles calculations suggest that they do not. Here we will show that the presence of this crossing probes a fundamental, but previously unexplored, property of the bulk electronic structure of Bi.

We will begin by pointing out that the mirror symmetry of the rhombohedral *A*7 structure leads to an additional topological structure of the energy bands which we refer to as a mirror Chern number. We will then show that the value of this integer in the topological insulator phase depends on the structure of the nearly degenerate L_s and L_a bands in pure Bi. We will identify a previously unexplored parameter in the $\mathbf{k} \cdot \mathbf{p}$ theory of Bi, which we refer to as the mirror chirality. We will show that the mirror chirality at the *L* point in Bi determines the value of the mirror Chern number in the topological insulator phase of $Bi_{1-x}Sb_x$.

We find that the value of the mirror chirality predicted by the Liu-Allen tight-binding model[26] disagrees with the value predicted by a more fundamental calculation by Golin.[34] This, combined with the disagreement with the surface-state experiments and first-principles calculations, suggests that the Liu-Allen tight-binding model has a subtle but topological error.

A. Mirror Chern number

The Dirac points in the hole pockets in our tight-binding calculation arise because the $\bar{\Sigma}_1$ and $\bar{\Sigma}_2$ bands cross on the line connecting $\bar{\Gamma}$ and $\bar{M}$ in Fig. 3(b). This crossing is protected by the invariance of the Hamiltonian under the mirror operation $\mathcal{M}(\hat{x}) = PC_2(\hat{x})$, which takes *x* to −*x*. $\bar{\Sigma}_1(\bar{\Sigma}_2)$ transform under different representations of $\mathcal{M}(\hat{x})$ with eigenvalues $+i(-i)$. This mirror symmetry implies that all the bulk electronic states in the plane $k_x = 0$ can be labeled with a mirror eigenvalue $\pm i$. Within this two-dimensional plane in momentum space, the occupied energy bands for each mirror eigenvalue will be associated with a *Chern invariant* $n_{\pm i}$. Time-reversal symmetry requires that $n_{+i} + n_{-i} = 0$, but the difference defines a nontrivial mirror Chern number,

$$n_{\mathcal{M}} = (n_{+i} - n_{-i})/2. \tag{5.1}$$

The situation is analogous to the quantum spin Hall state in graphene,[1,2] where the conservation of spin S_z in the two band model leads to the definition of a spin Chern number,[27] whose parity is related to the $\mathbb{Z}_2$ topological invariant. Since mirror symmetry is a physical lattice symmetry, the mirror Chern number is a fundamental characterization of a band structure. This distinguishes it from the spin Chern number, which is a property of a simplified model. Since the mirror Chern number relies on a spatial symmetry, it is a "weak" topological invariant in the sense discussed in Refs. 7 and

12. It loses its meaning in the presence of symmetry-breaking disorder. In principle there is a second mirror Chern number associated with the mirror-invariant plane $k_x = \pi/a$ in the Brillouin zone. For the band structures considered in this paper, this second invariant is zero, and will not be considered further.

The mirror Chern number determines how the surface states connect the valence and conduction bands along the line $q_x = 0$ between $\bar{\Gamma}$ to $\bar{M}$. To see this, consider the $\mathcal{M} = \pm i$ sectors independently. The bulk states with $k_x = 0$ are then analogous to a two-dimensional integer quantum Hall state with Hall conductivity $n_{\pm i}e^2/h$. The sign of $n_{\pm i}$ determines the direction of propagation of the edge states, which connect the valence and conduction bands. Thus, the sign of $n_{\mathcal{M}}$ determines whether the $\bar{\Sigma}_1$ band or the $\bar{\Sigma}_2$ band connects the valence and conduction bands between $\bar{\Gamma}$ and $\bar{M}$ (which we take to be in the $+\hat{y}$ direction). For $n_{\mathcal{M}} = +1$ (−1) we find that the $\bar{\Sigma}_1$ ($\bar{\Sigma}_2$) band crosses.

The predictions of the tight-binding model are more likely to be robust near $\bar{\Gamma}$ than near $\bar{M}$, because near $\bar{\Gamma}$ they are not sensitive to the detailed treatment of the small bulk energy gap at the *L* point. This is supported by the fact that the ordering of the $\bar{\Sigma}_1$ and $\bar{\Sigma}_2$ bands near $\bar{\Gamma}$ predicted by the tight-binding model (in which $\bar{\Sigma}_2$ emerges first) agrees with other calculations and experiment. Given this ordering near $\bar{\Gamma}$, the mirror Chern number determines whether or not the Σ_1 and Σ_2 bands have to cross. Referring to Fig. 3(c), if the mirror Chern number were to have the opposite sign, then the $\bar{\Sigma}_2$ band would connect to the conduction band rather than the $\bar{\Sigma}_1$ band, and the bands would not have to cross. Pure Bi is very close to the transition between the (0;000) and (1;111) phases. Therefore, it is likely that the presence of the crossing between $\bar{\Sigma}_1$ and $\bar{\Sigma}_2$ will be unaffected by the transition. Therefore, the sign of the $n_{\mathcal{M}}$ in the topological insulator phase of $Bi_{1-x}Sb_x$ should be correlated with the alternatives shown in Fig. 4, with $n_{\mathcal{M}} = +(-)1$ corresponding to Fig. 4(b) [Fig. 4(c)].

Since the valence band of pure Bi is in the trivial (0;000) topological class, pure Bi does not have surface states which connect the valence and conduction bands. Thus the mirror Chern number for the $k_x = 0$ plane of the valence band of pure Bi is $n_{\mathcal{M}} = 0$. The transition to the strong topological insulator in $Bi_{1-x}Sb_x$ occurs for small *x* because the L_s and L_a bands in pure Bi are nearly degenerate. At the transition to the topological insulator, the two bands cross and form a three-dimensional Dirac point at *L*. At this transition both the $\mathbb{Z}_2$ topological invariants $(\nu_0; \nu_1\nu_2\nu_3)$ and the mirror Chern number $n_{\mathcal{M}}$ change. The *change* $\Delta n_{\mathcal{M}}$ across this transition is an intrinsic property of this Dirac point. Thus the value of $n_{\mathcal{M}}$ in the topological insulator phase can be determined by studying the properties of this Dirac point. Since pure Bi is very close to this transition, this information can be extracted from the structure of the $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian for pure Bi in the vicinity of the *L* point.

In Sec. V B we will analyze the $\mathbf{k} \cdot \mathbf{p}$ theory and show that the value of $\Delta n_{\mathcal{M}}$ predicted by the Liu-Allen tight-binding model *disagrees* with the value predicted by an earlier

pseudopotential calculation by Golin.[34] This provides evidence that the crossing of the $\bar{\Sigma}_1$ and $\bar{\Sigma}_2$ bands is an artifact of the incorrect sign of $n_{\mathcal{M}}$ predicted by the tight-binding model.

B. k·p theory and the mirror chirality

The **k·p** analysis of Bi near the L point has a long history. Originally developed by Cohen and Blount[29] in 1960, the theory was given an particularly elegant formulation by Wolff,[30] who emphasized the similarity with the relativistic Dirac equation. This theory and its refinements[31–33] played an important role in the early development of band theory, and formed the framework for interpreting a large body of magnetic, transport, and optical data. In this section we point out a previously unexplored sign which characterizes this theory: the mirror chirality. We show that it is this sign which determines the sign of $n_{\mathcal{M}}$ in the topological insulator phase.

The four relevant states at the L point are denoted $[L_s, L_a] = [(L_6, L_5), (L_7, L_8)]$.[37] The two states comprising L_s and L_a are degenerate due to time-reversal symmetry. These states are distinguished by their symmetry under parity P (with eigenvalues $[(1,1),(-1,-1)]$), under the twofold rotation $C_2(\hat{x})$ (with eigenvalues $[(-i,i),(i,-i)]$) and under the mirror $\mathcal{M}(\hat{x}) = PC_2(\hat{x})$ (with eigenvalues $[(-i,i),(-i,i)]$). We have chosen the unconventional order of the states to simplify the mirror operator, which makes the connection with the mirror Chern number in Sec. V C the most transparent. In this basis the inversion, rotation, and mirror operators have the direct product form

$$P = \tau_z \otimes \mathbb{1},$$

$$C_2(\hat{x}) = -i\tau_z \otimes \mu_z,$$

$$\mathcal{M}(\hat{x}) = -i\mathbb{1} \otimes \mu_z, \qquad (5.2)$$

while the time-reversal operator can be chosen as

$$\Theta = i\mathbb{1} \otimes \mu_y K, \qquad (5.3)$$

where K is complex conjugation. $\vec{\mu}$ and $\vec{\tau}$ are Pauli matrices operating within and between the L_s and L_a blocks, and $\mathbb{1}$ is the identity matrix. In the following we will simplify the notation by omitting the $\otimes$ and the $\mathbb{1}$.

To first order in **k** the **k·p** Hamiltonian has the form

$$H(\mathbf{k}) = m\tau_z + k_x\Pi_x + k_y\Pi_y + k_z\Pi_z, \qquad (5.4)$$

where $E_G = 2m$ is the energy gap (positive for Bi) and Π_a are 4×4 matrices. Invariance of $H(\mathbf{k})$ under P and Θ requires $\{\Pi_a, P\} = \{\Pi_a, \Theta\} = 0$, and invariance under $\mathcal{M}(\hat{x})$ requires $\{\Pi_x, \mathcal{M}(\hat{x})\} = [\Pi_{y,z}, \mathcal{M}] = 0$. The allowed terms are thus

$$\Pi_x = t_1\tau_x\mu_x + t_2\tau_x\mu_y,$$

$$\Pi_y = u_{11}\tau_x\mu_z + u_{12}\tau_y,$$

$$\Pi_z = u_{21}\tau_x\mu_z + u_{22}\tau_y, \qquad (5.5)$$

where t_i and u_{ij} are real numbers. Equations (5.4) and (5.5) are equivalent to the **k·p** theory introduced by Cohen and

Blount,[29] who expressed the Hamiltonian in terms of the complex vectors **t** and **u**. These are related to our parameters via $\mathbf{t} = (t_1 + it_2)\hat{x}$ and $\mathbf{u} = (-u_{11} + iu_{12})\hat{y} + (-u_{21} + iu_{22})\hat{z}$. In the following it will be useful to express these in terms of three complex numbers $t = \hat{x}\cdot\mathbf{t}$ and $u^\pm = (\hat{y} \pm i\hat{z})\cdot\mathbf{u}$.

Equation (5.4) has a simpler form when expressed in terms of the principle axes in both momentum space and spin space. We thus perform a rotation $(k_y + ik_z) = e^{i\alpha}(k_y' + ik_z')$ along with a unitary transformation $|\psi\rangle = \exp[i\mu_z(\beta + \gamma\tau_z)]|\psi'\rangle$. These transformations have the effect of changing the phases, $t \to te^{i\beta}$ and $u^\pm \to u^\pm e^{-i(\gamma\pm\alpha)}$. For appropriately chosen α, β and γ, t and $\mp u^\pm$ can be made real and positive. The Hamiltonian then takes the diagonal form

$$H = m\tau_z + v_1 k_x \tau_x \mu_x + \eta v_2 k_y' \tau_x \mu_z + v_3 k_z' \tau_y, \qquad (5.6)$$

where

$$v_1 = |t|,$$

$$\eta v_2 = (|u^+| - |u^-|)/2,$$

$$v_3 = (|u^+| + |u^-|)/2. \qquad (5.7)$$

Here we have defined v_2 to be positive and introduced a previously unexplored quantity $\eta = \pm 1$, which is simply given by $\eta = \text{sgn}(\det[u_{ij}])$. η is a mirror chirality, which distinguishes two topologically distinct classes of Dirac Hamiltonians.

For a system with full rotational symmetry, η must be equal to $+1$. This can be seen by noting that the twofold rotation operator specifies the generator of continuous rotations about $\hat{x}$ via $C_2(\hat{x}) = \exp[-i\pi S_x]$. Since $C_2(\hat{x}) = -i\mu_z\tau_z$, this implies $S_x = \mu_z\tau_z/2$. When $\eta = -1$, Eq. (5.6) is *not* invariant under continuous rotations generated by S_x even when $v_2 = v_3$, since the spin and orbital degrees of freedom are rotated in opposite directions. The twofold rotational symmetry, however, remains intact. $\eta = +1$ corresponds to the behavior of a free electron and should be considered normal behavior. $\eta = -1$ is anomalous.

The sign of η is not ordinarily discussed in the **k·p** theory of Bi because it has no effect on the electronic dispersion $E(\mathbf{k})$, which depends only on $|v_a|$. η does, however, have a subtle effect in the presence of a magnetic field. A magnetic field in the $\hat{x}$ direction leads to a splitting of states according to their spin angular momentum S_x, which can be defined as above in terms of the twofold rotation operator $C_2(\hat{x})$. This defines a magnetic moment, which symmetry restricts to be either parallel or antiparallel to $\hat{x}$. The form of this magnetic moment is discussed in Refs. 30 and 31, and it is straightforward to show that $\vec{\mu} \propto \eta S_x\hat{x}$. This means η determines the *sign* of the g factor, which describes the relation between the magnetic moment and angular momentum. For $\eta = +1$ the sign is the same as that for a free electron, while for $\eta = -1$ the sign is opposite.

Unfortunately, this sign is difficult to probe experimentally. In addition to complications which arise due to the presence of three equivalent L points, measurement of the sign requires measurement of the spin angular momentum in addition to the change in energy with magnetic field. The selection rules discussed in Ref. 30 are unaffected by the

sign. We are not aware of any experiments on Bi which directly probe this sign.

C. Relation between mirror chirality, mirror Chern number, and surface states

We will now argue that the sign of η determines the sign of the mirror Chern number in the topological insulator phase of $Bi_{1-x}Sb_x$. This leads to an experimentally testable prediction regarding the crossing of the surface states. Thus, probing the surface states of the topological insulator may well be the best experimental method for determining this fundamental parameter of the $\mathbf{k} \cdot \mathbf{p}$ theory of Bi.

The connection between η and the mirror Chern number can be established by considering the mirror plane $k_x = 0$. H then decouples into two independent two band Hamiltonians for $\mathcal{M}(\hat{x}) = -i\mu_z = \pm i$ with the form

$$h = m\tau_z + sv_2 k_y' \tau_x + v_3 k_z' \tau_y, \tag{5.8}$$

where $s = \eta\mu_z$. $m = 0$ describes a transition where the Chern number $n_{-i\mu_z}$ changes. When m changes sign from negative to positive, $\Delta n_{-i\mu_z} = \eta\mu_z$. Thus, the change in the mirror Chern number,

$$\Delta n_{\mathcal{M}} = n_{\mathcal{M}}(m > 0) - n_{\mathcal{M}}(m < 0) = -\eta, \tag{5.9}$$

depends on the mirror chirality η. Since $n_{\mathcal{M}} = 0$ for Bi (with $m > 0$), we conclude that the topological insulator, with $m < 0$, has

$$n_{\mathcal{M}} = \eta. \tag{5.10}$$

$n_{\mathcal{M}}$ determines the direction of propagation of the $\bar{\Sigma}_1$ and $\bar{\Sigma}_2$ surface states along the mirror line $q_x = 0$. The direction of propagation of the surface states on the top surface which connect the valence and conduction bands can be determined by solving Eq. (5.8) with a z dependent mass $m(z) = m \, \mathrm{sgn}(z)$ with $m > 0$. The bound state at the surface has a wave function proportional to $\exp(-|mz|/v_3)$. The dispersion for the surface states on the top surface along $q_x = 0$ is

$$E(q_y) = -\eta\mu_z v q_y, \tag{5.11}$$

with $v > 0$. This means that the Σ_1 band, which has $\mu_z = -1$, propagates in the $+\eta\hat{y}$ direction, while the Σ_2 band, with $\mu_z = +1$, propagates in the $-\eta\hat{y}$ direction. Therefore, the surface state connecting the valence band to the conduction band which has the *positive* velocity in the $\hat{y}$ direction will be Σ_1 for $\eta = +1$ and Σ_2 for $\eta = -1$.

D. Comparison of tight-binding and pseudopotential models with experiment

In this subsection we show that the value of η predicted by the Liu-Allen tight-binding model[26] disagrees with that predicted by an early calculation by Golin.[34] Specifically, we find that the Liu-Allen model predicts the conventional value $\eta = 1$, while the Golin model predicts the anomalous value $\eta = -1$. We will then argue that the value of η can be extracted from the structure of the surface-state spectrum. The presently available spin polarized ARPES data on the Bi 111

TABLE III. Parameters of the $\mathbf{k} \cdot \mathbf{p}$ theory, Eq. (5.6), extracted from the pseudopotential model (Ref. 34) and the tight-binding model (Ref. 26).

	v_1 (eV Å)	v_2	v_3	η
Golin pseudopotential	4.16	1.37	7.01	-1
Liu-Allen tight-binding	5.89	0.92	9.67	$+1$

surface[22] provide indirect evidence that the mirror chirality has the anomalous value $\eta = -1$.

The $\mathbf{k} \cdot \mathbf{p}$ parameters can be determined by evaluating the matrix elements

$$\Pi_a^{ij} = \langle L_i | \hat{v}_a | L_j \rangle |_{\mathbf{k}=L}, \tag{5.12}$$

where $\hat{\mathbf{v}} = \nabla_{\mathbf{k}} \mathcal{H}(\mathbf{k})|_{\mathbf{k}=L}$ is determined by the Bloch Hamiltonian $\mathcal{H}(\mathbf{k})$. From this it follows that

$$t = \Pi_x^{57}, \tag{5.13}$$

$$u^{\pm} = -\Pi_y^{67} \mp i\Pi_z^{67}. \tag{5.14}$$

These matrix elements are listed in Table II of Golin's paper[34] (the relevant band is $j = j' = 3$). They may also be extracted from the Liu-Allen tight-binding model. In Table III we compare the values of v_1, v_2, v_3, and η computed from these matrix elements. The signs of η predicted by the two theories disagree. Since the parameters of the Liu-Allen model were simply fitted to reproduce the *energies* of the bands, there is no reason to expect that it gets η right. In contrast, Golin's calculation,[34] which is based on a pseudopotential approach, starts from more fundamental premise.

In Sec. V C we showed that provided there is only a single transition between pure Bi and the topological insulator phase of $Bi_{1-x}Sb_x$, the mirror chirality deduced from the pure Bi band structure determines the mirror Chern number in the topological insulator. This, in turn, determines the direction of propagation of the $\bar{\Sigma}_1$ and $\bar{\Sigma}_2$ states along the line $q_x = 0$. The surface-state structure predicted by the tight-binding model is shown in Fig. 3(c). The crossing of the Σ_1 band is consistent with $\eta = +1$. This crossing guarantees that there is a Dirac point enclosed by the hole pocket. This can be probed either by inverse photoemission or by spin polarized photoemission. In the latter case, the presence of the Dirac point would lead to a change in the sign of the spin on either side of the hole pocket. It will be interesting to experimentally determine this property for $Bi_{1-x}Sb_x$ using spin polarized ARPES.

Currently available spin polarized photoemission data on the 111 surface of pure Bi (Ref. 22) provide an indirect probe of η. Hole pockets are observed along the line from $\bar{\Gamma}$ to $\bar{M}$ in both $Bi_{1-x}Sb_x$ and pure Bi. Provided we make the plausible assumption that no additional level crossings occur near the transition to the topological insulator, then the presence or absence of Dirac points in the hole pockets should be the same on both sides of the transition. In Ref. 22, the spin in either side of the hole pocket was found to point in the same direction, which indicates that in pure Bi, the hole pockets do

not enclose a Dirac point. This conclusion was supported by first-principles surface-state calculations, which also find no crossing.[22] This suggests that in the alloy, it should be the Σ_2 band which connects the conduction and valence bands, which is consistent with $\eta = -1$.

It thus appears likely that the mirror chirality in Bi has the anomalous sign, $\eta = -1$. This conclusion contradicts the prediction of the tight-binding model, but it is supported by (1) the pseudopotential band structure of pure Bi and (2) the observed and calculated surface-state structures of pure Bi. Spin polarized ARPES experiments on the topological insulator $Bi_{1-x}Sb_x$ could more directly determine this sign by probing the mirror Chern number $n_{\mathcal{M}}$.

VI. CONCLUSION

In this paper we have analyzed the surface-state structure of the topological insulator $Bi_{1-x}Sb_x$. Using a simple tight-binding model based on Liu and Allen's[26] tight-binding parameterization, we confirmed that the surface states have the signature of the strong topological insulator by showing that the surface Fermi surface encloses an odd number of Dirac points. The tight-binding model also predicts that the surface is semimetallic, with an electron pocket centered on $\bar{\Gamma}$ along with six hole pockets.

Using general arguments based on inversion symmetry, we showed that the location of electron and hole pockets in the surface Brillouin zone is constrained by a quantity which we defined as the surface fermion parity. This quantity is determined by the parity invariants of the bulk band structure, and for a given surface it determines which surface TRIM are enclosed by an odd number of Fermi surface lines. This argument establishes a simple and direct connection between the bulk electronic structure and the surface electronic structure for crystals with inversion symmetry. Using this general principle, we predicted the structure of the surface states for several different faces of $Bi_{1-x}Sb_x$. For the 111 face, these predictions agree both with our surface-state calculations and with experiment. It will be interesting to test these predictions experimentally on other faces of $Bi_{1-x}Sb_x$.

Finally, we showed that the mirror symmetry present in the rhombohedral $A7$ lattice leads to additional topological structure in the bulk energy bands. We defined an integer mirror Chern number $n_{\mathcal{M}}$, whose value is nonzero in the topological insulator phase. The sign of $n_{\mathcal{M}}$ determines the direction of propagation of each of the surface states along the mirror plane, and thus determines which surface states connect the conduction and valence bands. We find that the crossing of the Σ_1 band predicted by the tight-binding model, which leads to a Dirac point in the hole pockets, disagrees with the natural extrapolation of experiments and first-principles calculations on pure Bi, which find no Dirac point in the hole pockets.

We traced this discrepancy to a previously unexplored property of the $\mathbf{k} \cdot \mathbf{p}$ band structure of pure Bi, which we defined as the mirror chirality η. We showed that η in pure Bi determines $n_{\mathcal{M}}$ in the topological insulator. Moreover, we showed that the Liu-Allen model predicts the conventional value $\eta = +1$, while an earlier pseudopotential calculation by

Golin[34] predicts the anomalous value $\eta = -1$. The latter value is consistent with the available experimental data on Bi, though the connection is rather indirect. A more direct test would be to directly measure the mirror chirality $n_{\mathcal{M}}$ in the topological insulator by probing the surface states with spin polarized ARPES.

It would be interesting to check that the value of η predicted by more accurate first-principles calculations of Bi agrees with the pseudopotential prediction. Since the tight-binding model was designed only to get the energies of the bands right, there is no reason to expect that it would get η right. It should be possible to come up with a new parametrization of the Liu-Allen model which would have $\eta = -1$. We expect that the surface states computed within this model would have band crossings which agree with experiment and first-principles calculations, though of course a quantitative description of the surface states requires an accurate description of the surface potential.

An important lesson to be learned from this paper is that in addition to time-reversal symmetry, spatial symmetries can play an important role in topologically constraining bulk and surface band structures. Our analysis of these symmetries has not been exhaustive. A complete theory of *topological band theory*, which accounts for the full point group symmetry of a crystal, is called for.

APPENDIX: SURFACE FERMION PARITY FROM BULK PARITY INVARIANTS

In this Appendix we show that for an inversion- and time-reversal-invariant crystal, the surface fermion number $N(\mathbf{q} = \Lambda_a)$ discussed in Sec. IV is an integer, whose parity is determined by the product of bulk parity invariants $\delta(\Gamma_{a1,2})$, which are products of parity eigenvalues given in Eqs. (2.1) and (3.5). The simple proof outlined here provides a direct connection between the topological structure of the surface states and the parity eigenvalues characterizing the bulk crystal.

The Bloch Hamiltonian $H(\Lambda_a, k_z)$ describes a parity- and time-reversal-invariant one-dimensional system. In the following we will suppress the dependence on Λ_a and consider a purely one-dimensional system. To determine the end charge N, we introduce the "cutting procedure" depicted in Fig. 7(a). We begin with a large but finite system with periodic boundary conditions. We then replace the hopping amplitudes t_i for all bonds that cross the cleavage plane $z=0$ by λt_i, where λ is real. Provided $z=0$ corresponds to an inversion plane, the one-dimensional Hamiltonian retains inversion and time-reversal symmetry for all λ. The fully cleaved crystal corresponds to $\lambda = 0$.
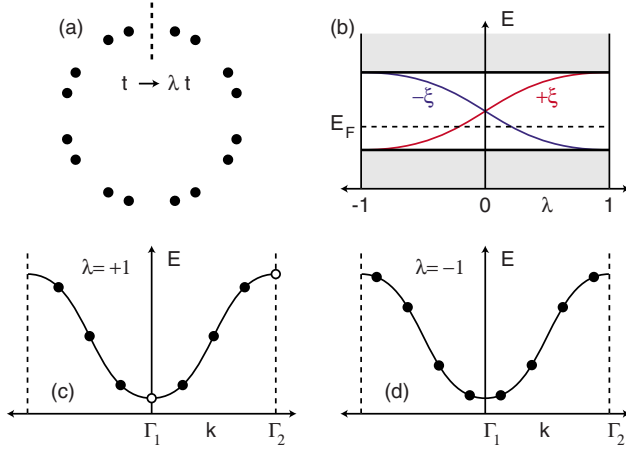
FIG. 7. (Color online) (a) A one-dimensional inversion-symmetric insulator cut at $z=0$ by replacing hopping amplitudes t across $z=0$ by λt. The fully cleaved crystal corresponds to $\lambda=0$. (b) Energy spectrum as a function of λ between -1 and 1. The conduction and valence bands exchange a Kramers pair of states with opposite parities. [(c) and (d)] The bulk energy levels at $\lambda=\pm 1$. For $\lambda=-1$ (d) every state at k has a partner at $-k$ with the same energy and opposite parity. For $\lambda=+1$ (c) the states at $k=\Gamma_1$ and $k=\Gamma_2$ are not paired.

For $\lambda=1$ the system is translationally invariant, so the excess charge near $z=0$ is $Q(\lambda=1)=0$. Since the insulator can have no bulk currents, the only way $Q(\lambda)$ can change is if a state localized near $z=0$ crosses the Fermi energy. Thus $Q(0)$ will be the difference between the number of states that cross E_F from above and from below for $\lambda \in [0,1]$. Kramers' theorem requires that every state is at least twofold degenerate, so the number of states crossing E_F will be an *even* integer. Since the charge will be divided evenly between the two sides, $N=Q(0)/2$ is an integer, which may be written as

$$N = \Delta N_+ - \Delta N_-, \qquad (A1)$$

where $\Delta N_\pm$ is the number of *Kramers pairs* that cross E_F from above or below.

We now relate the parity of N to the bulk parity eigenvalues. To this end it is useful to consider the evolution of the spectrum for $\lambda \in [-1,1]$ and to define

$$P(\lambda) = \prod_{E_{2\alpha}(\lambda) < E_F} \xi_{2\alpha} \qquad (A2)$$

as the product of the parities of all of the occupied states, where each Kramers pair $(\psi_{2\alpha}, \psi_{2\alpha-1})$ is included only once. This quantity is well defined because $\xi_{2\alpha} = \xi_{2\alpha-1}$. Our proof consists of two steps. We will first show that

$$P(1)P(-1) = (-1)^N. \qquad (A3)$$

We will then show that

$$P(1)P(-1) = \prod_{m=1}^{n_b} [-\xi_{2m}(\Gamma_1)\xi_{2m}(\Gamma_2)] \equiv \pi. \qquad (A4)$$

Here $\xi_{2m}(\Gamma_i)$ are the parity of the Bloch states in the mth Kramers degenerate band at the TRIM $k_z = \Gamma_i$, and again each Kramers pair is included only once. n_b is the number of occupied Kramers degenerate bands. Taken together, Eqs. (A3) and (A4) establish the relationship summarized by Eqs. (2.1) and (3.5) between the bulk parity eigenvalues and the surface fermion parity.

Equation (A3) follows from the symmetry of the end-state spectrum about $\lambda=0$. The Hamiltonian $H(-\lambda)$ differs from $H(\lambda)$ only by a phase twist of π across $z=0$. This twist can be spread over the entire circumference L by performing the gauge transformation

$$|\psi(-\lambda)\rangle = e^{i\pi z/L} |\tilde{\psi}(-\lambda)\rangle \qquad (A5)$$

for $0 < z < L$. When $L \to \infty$ the Hamiltonian for $|\tilde{\psi}(-\lambda)\rangle$ near $z=0$ becomes identical to $H(\lambda)$. Thus every bound state $|\psi_l(\lambda)\rangle$ satisfies $E_l(-\lambda) = E_l(\lambda)$. Since Eq. (A5) changes the parity, $|\psi_l(\lambda)\rangle$ and $|\psi_l(-\lambda)\rangle$ have opposite parities.

It follows that every Kramers pair that crosses the E_F at $\lambda_0 \in [1,0]$ has a partner with opposite parity that crosses E_F in the opposite direction at $-\lambda_0$ as shown in Fig. 7(b). Thus between $\lambda=1$ and $\lambda=-1$, the conduction and valence bands exchange two Kramers pairs with opposite parities, leading to a change in the relative sign between $P(1)$ and $P(-1)$. We conclude that $P(1)P(-1) = (-1)^{\Delta N_+ + \Delta N_-}$, which leads directly to Eq. (A3).

Equation (A4) follows from a consideration of the parities of the Bloch wave functions. Consider first the simplest case where there is a single Kramers degenerate occupied band, as shown in Figs. 7(c) and 7(d). At $\lambda=1$ the single-particle states are labeled by momentum $k_z = 2m\pi/L$ with $m = -M/2 + 1, \ldots, M/2$, where M is the number of unit cells. At the two TRIM $\Gamma_1 = 0$ and $\Gamma_2 = M\pi/L$, the parity eigenvalues are $\xi(\Gamma_{1,2})$. Every other k_z has a partner $-k_z$, and even- and odd-parity combinations of the two can be formed. The $M/2-1$ $(k_z, -k_z)$ pairs thus each contribute -1 to the product in Eq. (A2). Therefore,

$$P(1) = (-1)^{M/2-1} \xi(\Gamma_1)\xi(\Gamma_2). \qquad (A6)$$

For $\lambda=-1$ gauge transformation (A5) leads to a periodic Hamiltonian identical to $H(1)$, but with momenta shifted by π/L, as shown in Fig. 7(d). Thus all the momenta are paired, so that

$$P(-1) = (-1)^{M/2}. \qquad (A7)$$

Combining Eqs. (A6) and (A7) leads directly to Eq. (A4), which is straightforwardly generalized to the case of n_b Kramers degenerate bands.

[1] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 226801 (2005).

[2] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 146802 (2005).

[3] B. A. Bernevig and S. C. Zhang, Phys. Rev. Lett. **96**, 106802 (2006).

[4] R. Roy, arXiv:cond-mat/0604211 (unpublished).

[5] J. E. Moore and L. Balents, Phys. Rev. B **75**, 121306(R) (2007).

[6] R. Roy, arXiv:cond-mat/0607531 (unpublished).

[7] L. Fu, C. L. Kane, and E. J. Mele, Phys. Rev. Lett. **98**, 106803 (2007).

[8] B. A. Bernevig, T. Hughes, and S. C. Zhang, Science **314**, 1757 (2006).

[9] M. König, S. Wiedmann, C. Brüne, A. Roth, H. Buhmann, L. Molenkamp, X. L. Qi, and S. C. Zhang, Science **318**, 766 (2007).

[10] L. Fu and C. L. Kane, Phys. Rev. Lett. **100**, 096407 (2008).

[11] X. L. Qi, T. Hughes, and S. C. Zhang, arXiv:0802.3537 (unpublished).

[12] L. Fu and C. L. Kane, Phys. Rev. B **76**, 045302 (2007).

[13] For a review, see Ph. Hofmann, Prog. Surf. Sci. **81**, 191 (2006).

[14] F. Patthey, W. D. Schneider, and H. Micklitz, Phys. Rev. B **49**, 11293 (1994).

[15] S. Agergaard, C. Sondergaard, H. Li, M. B. Nielsen, S. V. Hoffmann, Z. Li and Ph. Hofmann, New J. Phys. **3**, 15 (2001).

[16] C. R. Ast and H. Höchst, Phys. Rev. Lett. **87**, 177602 (2001).

[17] C. R. Ast and H. Höchst, Phys. Rev. B **66**, 125103 (2002).

[18] J. I. Pascual, G. Bihlmayer, Yu. M. Koroteev, H. P. Rust, G. Ceballos, M. Hansmann, K. Horn, E. V. Chulkov, S. Blügel, P. M. Echenique, and Ph. Hofmann, Phys. Rev. Lett. **93**, 196802 (2004).

[19] Yu. M. Koroteev, G. Bihlmayer, J. E. Gayone, E. V. Chulkov, S. Blügel, P. M. Echenique, and Ph. Hofmann, Phys. Rev. Lett. **93**, 046403 (2004).

[20] Ph. Hofmann, J. E. Gayone, G. Bihlmayer, Y. M. Koroteev, and E. V. Chulkov, Phys. Rev. B **71**, 195413 (2005).

[21] T. Hirahara, T. Nagao, I. Matsuda, G. Bihlmayer, E. V. Chulkov, Y. M. Koroteev, P. M. Echenique, M. Saito, and S. Hasegawa, Phys. Rev. Lett. **97**, 146803 (2006).

[22] T. Hirahara, K. Miyamoto, I. Matsuda, T. Kadono, A. Kimura, T. Nagao, G. Bihlmayer, E. V. Chulkov, S. Qiao, K. Shimada, H. Namatame, M. Taniguchi, and S. Hasegawa, Phys. Rev. B **76**, 153305 (2007).

[23] H. Höchst and S. A. Gorovikov, J. Electron Spectrosc. Relat. Phenom. **144**, 351 (2005).

[24] D. Hsieh, D. Qian, L. Wray, Y. Xia, Y. Hor, R. J. Cava, and M. Z. Hasan, Nature (London) **452**, 970 (2008).

[25] Y. M. Koroteev, G. Bihlmayer, E. V. Chulkov, and S. Blügel, Phys. Rev. B **77**, 045428 (2008).

[26] Y. Liu and R. E. Allen, Phys. Rev. B **52**, 1566 (1995).

[27] D. N. Sheng, Z. Y. Weng, L. Sheng, and F. D. M. Haldane, Phys. Rev. Lett. **97**, 036808 (2006).

[28] L. Fu and C. L. Kane, Phys. Rev. B **74**, 195312 (2006).

[29] M. H. Cohen and E. I. Blount, Philos. Mag. **5**, 115 (1960).

[30] P. A. Wolff, J. Phys. Chem. Solids **25**, 1057 (1964).

[31] G. E. Smith, G. A. Baraff, and J. M. Rowell, Phys. Rev. **135**, A1118 (1964).

[32] G. A. Baraff, Phys. Rev. **137**, A842 (1965).

[33] F. A. Buot and J. W. McClure, Phys. Rev. B **6**, 4525 (1972).

[34] S. Golin, Phys. Rev. **166**, 643 (1968).

[35] B. Lenoir, M. Cassart, J.-P. Michenaud, H. Scherrer, and S. Scherrer, J. Phys. Chem. Solids **57**, 89 (1996).

[36] B. Lenoir, A. Dauscher, X. Devaux, R. Martin-Lopez, Yu. I. Ravich, H. Scherrer, and S. Scherrer, Proceedings of the 15th International Conference on Thermoelectrics, Pasadena, CA (IEEE, New York, 1996), p. 1.

[37] L. M. Falicov and S. Golin, Phys. Rev. **137**, A871 (1965).

[38] D. H. Lee and J. D. Joannopoulos, Phys. Rev. B **23**, 4988 (1981).

[39] F. Claro, Phys. Rev. B **17**, 699 (1978).

Critical behavior of a point contact in a quantum spin Hall insulator

Jeffrey C. Y. Teo and C. L. Kane

Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

We study a quantum point contact in a quantum spin Hall insulator. It has recently been shown that the Luttinger liquid theory of such a structure maps to the theory of a weak link in a Luttinger liquid with spin with Luttinger liquid parameters $g_\rho = 1/g_\sigma = g < 1$. We show that for weak interactions, $1/2 < g < 1$, the pinch-off of the point contact as a function of gate voltage is controlled by a novel quantum critical point, which is a realization of a nontrivial intermediate fixed point found previously in the Luttinger liquid model with spin. We predict that the dependence of the conductance on gate voltage near the pinch-off transition for different temperatures collapses onto a universal curve described by a crossover scaling function associated with that fixed point. We compute the conductance and critical exponents of the critical point as well as the universal scaling function in solvable limits, which include $g = 1 - \epsilon$, $g = 1/2 + \epsilon$, and $g = 1/\sqrt{3}$. These results, along with a general scaling analysis, provide an overall picture of the critical behavior as a function of g. In addition, we analyze the structure of the four-terminal conductance of the point contact in the weak tunneling and weak backscattering limits. We find that different components of the conductance can have different temperature dependences. In particular, we identify a skew conductance G_{XY}, which we predict vanishes as T^γ with $\gamma \geq 2$. This behavior is a direct consequence of the unique edge state structure of the quantum spin Hall insulator. Finally, we show that for strong interactions, $g < 1/2$, the presence of spin nonconserving spin-orbit interactions leads to a novel time-reversal-symmetry breaking insulating phase. In this phase, the transport is carried by spinless chargons and chargeless spinons. These lead to nontrivial correlations in the low frequency shot noise. Implications for experiments on HgCdTe quantum well structures will be discussed.

I. INTRODUCTION

A quantum spin Hall insulator (QSHI) is a time-reversal invariant two-dimensional electronic phase which has a bulk energy gap generated by the spin-orbit interaction.[1,2] It has a topological order[3] which requires the presence of gapless edge states similar to those that occur in the integer quantum Hall effect. In the simplest version, the QSHI can be understood as two time reversed copies of the integer quantum Hall state[4] for up and down spins. The edge states, which propagate in opposite directions for the two spins, form a unique one-dimensional (1D) system in which elastic backscattering is forbidden by time-reversal symmetry.[1] This state occurs in HgCdTe quantum well structures,[5] and experiments have verified the basic features of the edge states, including the Landauer conductance[6] $2e^2/h$, as well as the nonlocality of the edge state transport.[7]

In the presence of electron interactions, the edge states form a Luttinger liquid.[8–14] For strong interactions (when the Luttinger liquid parameter $g < 3/8$) random two particle backscattering processes destabilize the edge states, leading to an Anderson localized phase. For $g > 3/8$ (or a sufficiently clean system), however, one expects the characteristic power-law behavior for tunneling of a Luttinger liquid.

A powerful tool for probing edge state transport experimentally is to make a quantum point contact. As depicted in Figs. 1(a) and 1(b), a gate voltage controls the coupling between edge states on either side of a Hall bar as the point contact is pinched off. Recently, the point contact problem for a QSHI has been studied.[10,11] Hou *et al.*[10] made the interesting observation that the QSHI problem maps to an earlier studied model[15,16] of a weak link in a spinful Luttinger

liquid (SLL), in which the charge and spin Luttinger parameters are given by $g_\rho = g$ and $g_\sigma = 1/g$.[17] For sufficiently strong interactions ($g < 1/2$) they found that the simple perfectly transmitting and perfectly reflecting phases are both unstable. They showed that as long as spin is conserved at the junction the low energy behavior is dominated by a non-
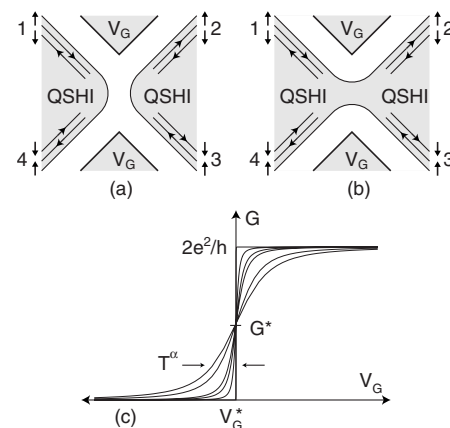
FIG. 1. A quantum point contact in a QSHI, controlled by a gate voltage V_G. In (a) $V_G < V_G^*$ and the point contact is pinched off. The spin filtered edge states are perfectly reflected. In (b) $V_G > V_G^*$ and the point contact is open. The edge states are perfectly transmitted. In (c) we plot the conductance (later defined as G_{XX}) as a function of V_G for different temperatures. As the temperature is lowered, the pinch-off curve sharpens up with a width T^α. The curves cross at a critical conductance G^*, and the shape of the curve has the universal scaling form (1.1). The plotted curves are based on Eq. (3.9), valid for $g = 1 - \epsilon$, which is computed in Sec. III C.

trivial "mixed" fixed point of the SLL, in which charge is reflected but spin is perfectly transmitted. This charge insulator/spin conductor (IC) phase leads to a novel structure in the four-terminal conductance of the point contact.

In this paper, we will focus on the QSHI point contact for weaker interactions when $1/2 < g < 1$. In this regime the open limit (or weak backscattering, "small v") and the pinched-off limit (or weak tunneling, "small t") are *both* stable perturbatively. This is different from the behavior in an ordinary Luttinger liquid[15,16,18] or a fractional quantum Hall point contact.[19,20] In those cases the perfectly transmitting limit is unstable for $g < 1$. Weak backscattering is relevant and grows at low energy, leading to a crossover to the stable perfectly reflecting fixed point. The fact that both the small v and the small t limits are stable for the QSHI point contact means that there must be an intermediate unstable fixed point which separates the flows to the two limits. This unstable fixed point describes a quantum critical point where the point contact switches on as a function of the pinch-off gate voltage. We will argue that in the limit of zero temperature the point contact switches on abruptly as a function of gate voltage V_G, with conductances $G = 0$ for $V_G < V_G^*$ and $G = 2e^2/h$ for $V_G > V_G^*$. At finite but low temperature T, the shape of the pinch-off curve $G(V_G, T)$ is controlled by the crossover between the unstable and stable fixed points and is described by a universal crossover scaling function,

$$\lim_{\Delta V_G, T \to 0} G(V_G, T) = \frac{2e^2}{h} \mathcal{G}_g\left(c \frac{\Delta V_G}{T^{\alpha_g}}\right). \quad (1.1)$$

Here $\Delta V_G = V_G - V_G^*$ and c is a nonuniversal constant. α_g is a critical exponent describing the unstable intermediate fixed point. $\mathcal{G}_g(X)$ is a universal function which crosses over between 0 and 1 as a function of X. α_g and $\mathcal{G}_g(X)$ are completely determined by the Luttinger liquid parameter g. This behavior means that as temperature is lowered, the pinch-off curve as a function of V_G sharpens up with a characteristic width which vanishes as T^{α_g}, as shown schematically in Fig. 1(c). The curves at different low temperatures cross at G_g^* = $\mathcal{G}_g(0)$, the conductance of the critical point. Equation (1.1) predicts that data from different temperatures can be rescaled to lie on the same universal curve.

The crossover scaling function $\mathcal{G}_g(X)$ is similar to the scaling function that controls the line shape of resonances in a Luttinger liquid[16,21] and in a fractional quantum Hall point contact.[19] That scaling function was computed exactly for all g by Fendley *et al.*[22] using the thermodynamic Bethe ansatz. That problem, however, was simpler than ours because the critical point occurs at the weak backscattering limit, which is described by a boundary conformal field theory with a trivial boundary condition.[23] The intermediate fixed point relevant to our problem has no such simple description. Thus, even the critical point properties α_g and G_g^* (which were simple for the resonance problem) are highly nontrivial to determine.

Intermediate fixed points in Luttinger liquid problems were first discussed in Refs. 15 and 16 in the context of SLLs. However, for that problem they occur in a rather unphysical region of parameter space $g_\sigma > 2$ because spin rota-

tional invariance requires $g_\sigma = 1$. The QSHI point contact provides a physically viable system to directly probe these nontrivial fixed points.

The existence of the intermediate fixed points can be inferred from the stability of the simple perfectly transmitting or reflecting fixed points.[15,16] However their properties are difficult to compute, and a general characterization of these critical points remains an unsolved problem in conformal field theory.[24] Two approaches have been used to study their properties. In Ref. 16, a perturbative approach was introduced which applies when the Luttinger parameters are close to their critical values g_σ^* and g_ρ^*, where the simple fixed points become unstable. (For instance, for the weak backscattering limit, $g_\rho^* = 1/2$, $g_\sigma^* = 3/2$.) For $g_{\rho,\sigma} = g_{\rho,\sigma}^* - \epsilon$, the fixed point is accessible in perturbation theory about the simple fixed point, and its properties can be computed in a manner analogous to the ϵ expansion in statistical mechanics.

An alternative approach is to map the theory for specific values of g_ρ and g_σ onto solvable models. In Ref. 25, Yi and Kane recast the Luttinger liquid barrier problem as a problem of quantum Brownian motion (QBM) in a two-dimensional periodic potential. When $g_\rho = 1/3$, $g_\sigma = 1$ and the potential has minima with a honeycomb lattice symmetry, a stable intermediate fixed point which occurs in that problem was identified with that of the three-channel Kondo problem. This, in turn, is related to the solvable $SU(2)_3$ Wess-Zumino-Witten model,[26] allowing for a complete characterization of the fixed point. This idea was further developed by Affleck *et al.*,[24] who provided a more general characterization of the fixed point in terms of the boundary conformal field theory of the three-state Pott's model. For $g_\rho = 1/\sqrt{3}$ and $g_\sigma = \sqrt{3}$ the QBM model with triangular lattice symmetry has an unstable intermediate fixed point, which we will see is related to the fixed point of the QSHI problem. In Ref. 25 symmetry arguments were exploited to determine the critical conductance G^* in that case.

In this paper we will compute α_g and $\mathcal{G}_g(X)$ (along with a multiterminal generalization of the conductance) in three solvable limits:

(i) For $g = 1 - \epsilon$, we will perform an expansion for weak electron interactions. For noninteracting electrons the point contact can be characterized in terms of a scattering matrix S_{ij} for arbitrary transmission. Weak interactions lead to a logarithmic renormalization of S_{ij}. Following the method developed by Matveev *et al.*,[27] this allows $\mathcal{G}_g(X)$ and α_g to be calculated exactly in the limit $g \to 1$.

(ii) For $g = 1/2 + \epsilon$, we find that the intermediate fixed point approaches the charge insulator/spin conductor fixed point, allowing for a perturbative calculation of the fixed point properties G_g^* and α_g to leading order in ϵ. Moreover, for $g = 1/2$ the Luttinger liquid theory can be fermionized, which allows the full crossover function $\mathcal{G}_g(X)$ to be determined in that limit.

(iii) For $g = 1/\sqrt{3}$, the self-duality argument developed in Ref. 25 allows us to compute the fixed point conductance G^* exactly.

These three results, along with the general scaling analysis, provide an overall picture of the critical behavior of the QSHI point contact as a function of g.

In addition to the analysis of the pinch-off transition discussed above, we will touch on two other issues in this paper.

First, we will introduce a convenient parametrization of the four-terminal conductance as a 3×3 conductance matrix. In this form symmetry constraints on the conductance are reflected in a natural way. Moreover, we will predict that different components of the conductance matrix have different temperature dependences at the low temperature fixed points. In particular, we will introduce a "skew" conductance G_{XY}, which is predicted to vanish as T^γ with $\gamma \geq 2$. For noninteracting electrons we will show that $G_{XY} = 0$ and for weak interactions $\gamma = 2$. This behavior is a direct consequence of the spin filtered nature of the edge states and does not occur in a generic four-terminal conductance device. It is thus a powerful diagnostic for the edge states.

Second, we will examine the role of spin-orbit terms at the point contact which respect time-reversal symmetry but violate spin conservation. For $g > 1/2$ we will provide evidence that such terms are *irrelevant* at the intermediate critical fixed point, so that they are unimportant for the critical behavior of the point contact. However, for $g < 1/2$, such terms are relevant. Hou *et al.*[10] pointed out that these terms are relevant perturbations at the charge insulator/spin conductor fixed point for $g < 1/2$, but they did not identify the stable phase to which the system flows at low energy. We will argue that the system flows to a time-reversal-symmetry breaking insulating state in which the *four-terminal* conductance $G_{ij} = 0$. Since spin-orbit interaction terms will generically be present in a point contact, the true low energy behavior of a point contact will be described by this phase. An interesting consequence of the broken time-reversal symmetry of this phase is that the weak tunneling processes which dominate the conductance at low but finite temperature are not electron tunneling processes. Rather, they involve the tunneling of neutral spinons and spinless chargons. This has nontrivial implications for four-terminal noise correlation measurements. A related effect has been predicted by Maciejko *et al.*[12] for the insulating state of a single impurity on a single edge of a QSHI. This insulating state, however, requires stronger electron-electron interactions. It occurs in the regime $g < 1/4$, where weak disorder already leads to Anderson localization.

This paper is organized as follows. In Sec. II we discuss our model and analyze five stable phases. In addition to the simple fixed points, where charge and spin are either perfectly reflected or perfectly transmitted, we discuss the time-reversal-symmetry breaking insulating phase which occurs for strong interactions with spin orbit. In Sec. III we discuss the critical behavior of the conductance at the pinch-off transition. We will begin in Sec. III A with a general discussion of the scaling theory and phase diagram along with a summary of our results. Readers who are not interested in the detailed calculations can go directly to this section. In the following sections we describe the calculations for $g = 1/\sqrt{3}$, $g = 1 - \epsilon$, and $g = 1/2 + \epsilon$ in detail. In Sec. IV we conclude with a discussion of experimental and theoretical issues raised by this work. In the Appendix we describe our parametrization of the four-terminal conductance and show that in this representation symmetry constraints have a simple form.

II. MODEL AND STABLE PHASES

In this section we will describe the Luttinger liquid theory of the QSHI point contact. We will begin in Sec. II A by

describing the Luttinger liquid model first for a single edge and then relating the four edges to the theory of the SLL. We then discuss the four-terminal conductance. In Sec. II B we describe the simple limits of our model which correspond to stable phases. The simplest limits are the perfect transmission limit or charge conductor/spin conductor (CC) and the perfect reflection limit or charge insulator/spin insulator (II). In addition we will discuss the "mixed" phases, including the IC and the charge conductor/spin insulator (CI).

For most of this section we will assume that spin is conserved. While spin nonconserving spin-orbit interactions are allowed and will generically be present we will argue that they are irrelevant for the fixed points and crossovers of physical interest. An exception to this, however, occurs for strong interactions when $g < 1/2$. This will be discussed in Sec. II B 5, where we will show that there are relevant spin-orbit terms which destabilize the CC, II, and IC phases. We will argue that these perturbations flow to a different low temperature phase, which we identify as a time-reversal-symmetry breaking insulator (TBI). In that section we will explore the transport properties of that state.

Much of the theory presented in this section is contained either explicitly or implicitly in the work of Hou *et al.*,[10] as well as in Refs. 11, 15, and 16. We include it here to establish our notation and to make our discussion self-contained. We will highlight, however, three results of this section which are original to this work. They include (1) our analysis of the four-terminal conductance, which predicts that different components of the conductance matrix have different temperature dependences. In particular, we find that the skew conductance G_{XY} vanishes at low temperature as T^γ with $\gamma \geq 2$. (2) In Sec. II B 5 we introduce the TBI phase discussed above. (3) We introduce a perturbative analysis of the IC and CI phases in Secs. II B 3 and II B 4. While this was partially discussed in Ref. 16, we will show that a full analysis requires the introduction of a pseudospin degree of freedom in the perturbation theory. This new pseudospin does not affect the lowest order stability analysis of the IC phase, but it will prove crucial for the second-order renormalization group flows, which will be used in the ϵ expansion in Sec. III D.

A. Model

The edge states on the four edges in Figs. 1(a) and 1(b) emanating from the point contact may be described by the Hamiltonian

$$H_0 = \sum_{i=1}^{4} \int_0^\infty dx_i \mathcal{H}_0^i, \tag{2.1}$$

with

$$\mathcal{H}_0^i = iv_0(\psi_{i,\text{in}}^\dagger \partial_x \psi_{i,\text{in}} - \psi_{i,\text{out}}^\dagger \partial_x \psi_{i,\text{out}}) + u_2 \psi_{i,\text{in}}^\dagger \psi_{i,\text{in}} \psi_{i,\text{out}}^\dagger \psi_{i,\text{out}}$$

$$+ \frac{1}{2} u_4 [(\psi_{i,\text{in}}^\dagger \psi_{i,\text{in}})^2 + (\psi_{i,\text{out}}^\dagger \psi_{i,\text{out}})^2]. \tag{2.2}$$

Here $\psi_{i,\text{in}}$ and $\psi_{i,\text{out}}$ are a time reversed pair of fermion operators with opposite spin which propagate toward and away from the junction. v_0 is the bare Fermi velocity and u is

electron interaction strength. u_2 and u_4 are forward scattering interaction parameters. The boundary condition on the fermions at $x=0$ is determined by the transmission of the point contact and will be discussed in various limits below.

1. Bosonization of a single edge

We first consider the Luttinger liquid theory for a single edge. We thus bosonize according to

$$\psi_{i,a} = \frac{1}{\sqrt{2\pi x_c}} e^{i\phi_{i,a}}, \qquad (2.3)$$

where $a=\text{in,out}$ and x_c is a short distance cutoff. $\psi_{i,a}$ obey the Kac Moody commutation algebra,

$$[\phi_{i,a}(x), \phi_{j,b}(y)] = i\pi\delta_{ij}\tau^z_{ab}\,\text{sgn}(x-y). \qquad (2.4)$$

Then,

$$\mathcal{H}_0^i = \frac{v_0}{4\pi}\{(1+\lambda_4)[(\partial_x\phi_{i,\text{in}})^2 + (\partial_x\phi_{i,\text{out}})^2] - 2\lambda_2\partial_x\phi_{i,\text{in}}\partial_x\phi_{i,\text{out}}\}, \qquad (2.5)$$

where $\lambda_i = u_i/(2\pi v_0)$. Changing variables

$$\begin{pmatrix}\phi_{i,\text{in}} \\ \phi_{i,\text{out}}\end{pmatrix} = \frac{1}{2g}\begin{pmatrix}1+g & 1-g \\ 1-g & 1+g\end{pmatrix}\begin{pmatrix}\tilde{\phi}_{i,\text{in}} \\ \tilde{\phi}_{i,\text{out}}\end{pmatrix} \qquad (2.6)$$

transforms Eq. (2.5) into a theory of decoupled chiral bosons,

$$\mathcal{H}_0 = \frac{v}{4\pi g}[(\partial_x\tilde{\phi}_{i,\text{in}})^2 + (\partial_x\tilde{\phi}_{i,\text{out}})^2], \qquad (2.7)$$

where $\tilde{\phi}_{i,a}$ obey

$$[\tilde{\phi}_{i,a}(x), \tilde{\phi}_{j,b}(y)] = i\pi g\,\delta_{ij}\tau^z_{ab}\,\text{sgn}(x-y). \qquad (2.8)$$

Here $v = v_0\sqrt{(1+\lambda_4)^2-\lambda_2^2}$ and

$$g = \sqrt{\frac{1+\lambda_4-\lambda_2}{1+\lambda_4+\lambda_2}}. \qquad (2.9)$$

The Luttinger liquid parameter g determines the power-law exponents for various quantities. For instance, the tunneling density of states scales as $\rho(E) \propto E^{(g+1/g)/2-1}$.

2. Mapping to spinful Luttinger liquid

Consider an open point contact in a Hall bar geometry with edge states on the top and bottom edges which continuously connect leads 1 and 2 and leads 3 and 4. We then define left and right moving fields with spin $\uparrow,\downarrow$ as

$$\phi_{R\uparrow} = \phi_{1,\text{in}}(-x)\theta(-x) + \phi_{2,\text{out}}(x)\theta(x),$$

$$\phi_{L\downarrow} = \phi_{2,\text{in}}(x)\theta(x) + \phi_{1,\text{out}}(-x)\theta(-x),$$

$$\phi_{L\uparrow} = \phi_{3,\text{in}}(x)\theta(x) + \phi_{4,\text{out}}(-x)\theta(-x),$$

$$\phi_{R\downarrow} = \phi_{4,\text{in}}(-x)\theta(-x) + \phi_{3,\text{out}}(x)\theta(x). \qquad (2.10)$$

It is then useful to define sum and difference fields as

TABLE I. The effect of discrete symmetry operations on the boson fields θ_ρ and θ_σ.

O	$\Theta O \Theta^{-1}$	$\mathcal{M}_X O \mathcal{M}_X^{-1}$	$\mathcal{M}_Y O \mathcal{M}_Y^{-1}$
θ_ρ	θ_ρ	$-\theta_\rho$	θ_ρ
φ_ρ	$-\varphi_\rho$	φ_ρ	φ_ρ
θ_σ	$-\theta_\sigma$	θ_σ	$-\theta_\sigma$
φ_σ	$\varphi_\sigma + \pi$	$-\varphi_\sigma$	$-\varphi_\sigma$

$$\phi_{a\sigma} = \frac{1}{2}(\varphi_\rho + \sigma\varphi_\sigma + a\theta_\rho + a\sigma\theta_\sigma), \qquad (2.11)$$

where $a=R,L=+,-$ and $\sigma=\uparrow,\downarrow=+,-$. Then, θ_α and ϕ_α obey

$$[\theta_\alpha(x), \phi_\beta(y)] = 2\pi i\delta_{\alpha\beta}\theta(x-y), \qquad (2.12)$$

and Eqs. (2.3) and (2.5) become[10]

$$H_0 = \int_{-\infty}^{\infty} dx \sum_{a=\sigma,\rho} \frac{v}{4\pi}\left[g_a(\partial_x\varphi_a)^2 + \frac{1}{g_a}(\partial_x\theta_a)^2\right], \qquad (2.13)$$

where

$$g_\rho = g, \quad g_\sigma = 1/g, \qquad (2.14)$$

and g and v are given in Sec. II A 1.

It is useful to list the effect of symmetry operations on the charge-spin variables because symmetries constrain the allowed tunneling operators. Charge conservation leads to gauge invariance under the transformation $\phi_\rho \to \phi_\rho + \delta_\rho$. The conservation of spin S_z leads to invariance under $\phi_\sigma \to \phi_\sigma + \delta_\sigma$. The effects of time-reversal and mirror symmetries are shown in Table I. Time-reversal symmetry is specified by the operation $\Theta\psi_{a\sigma}\Theta^{-1}=i\sigma\psi_{\bar{a}\bar{\sigma}}$. The mirror $\mathcal{M}_X$ interchanges leads $14\leftrightarrow23$ while $\mathcal{M}_Y$ interchanges leads $12\leftrightarrow34$.

3. Four-terminal conductance

The central measurable quantity is the four-terminal conductance, defined by

$$I_i = \sum_j G_{ij}V_j, \qquad (2.15)$$

where I_i is the current flowing into lead i. G_{ij} is in general characterizing by nine independent parameters. In the Appendix we introduce a convenient representation for these parameters, which simplifies the representation of symmetry constraints. Here we will summarize the key points of that analysis.

The presence of both time-reversal symmetry and spin conservation considerably simplifies the conductance. It is characterized by *three* independent conductances,

$$\begin{pmatrix}I_X \\ I_Y\end{pmatrix} = \begin{pmatrix}G_{XX} & G_{XY} \\ G_{YX} & G_{YY}\end{pmatrix}\begin{pmatrix}V_X \\ V_Y\end{pmatrix}. \qquad (2.16)$$

Here $I_X=I_1+I_4$ is the current flowing from left to right in Fig. 1, while $I_Y=I_1+I_2$ is the current flowing from top to bottom. Similarly, V_X is a voltage biasing lead (2.14) relative to Eq.

(2.23) and V_Y biases lead (2.12) relative to Eq. (2.34). G_{XX} is thus the *two-terminal* conductance measured horizontally, while G_{YY} is the two-terminal conductance measured vertically. $G_{XY} = G_{YX}$ is a "skew conductance," which vanishes in the presence of mirror symmetry. Given these three parameters, the full four-terminal conductance matrix G_{ij} can be constructed using Eq. (A6).

A second consequence of spin conservation is the quantization of a particular combination of G_{ij}. In particular, in the Appendix we define a third current $I_Z = I_1 + I_3$ and a third voltage V_Z which biases lead (2.13) relative to Eq. (2.24). Spin conservation then requires

$$I_Z = G_{ZZ}V_Z, \quad (2.17)$$

with

$$G_{ZZ} = 2\frac{e^2}{h}. \quad (2.18)$$

Since spin nonconserving spin-orbit terms are allowed, spin conservation will not be generically present in the microscopic Hamiltonian of the junction. Nonetheless, we will argue that the low temperature fixed points possess an *emergent* spin conservation, as well as mirror symmetry, so that Eq. (2.18) should hold, albeit with corrections which vanish as a power of temperature.

B. Stable phases

In this section we describe various stable fixed points which admit simple descriptions using bosonization. We will first focus on the limit in which spin is conserved at the junction. There are then four simple fixed points.[15,16] These include the perfectly transmitting (CC) limit, in which both charge and spin conduct, and the perfectly reflecting limit (II), in which both charge and spin are insulating. The mixed fixed points, denoted IC (CI), are perfectly reflecting for charge (spin) and perfectly transmitting for spin (charge).

In the presence of spin nonconserving spin-orbit terms (which preserve time-reversal symmetry) an additional fixed point is possible in which time-reversal symmetry is spontaneously broken. We will see that in the presence of spin-orbit terms this TBI phase is the stable phase when $g < 1/2$.

1. Weak backscattering (CC) limit

We first consider the limit where the point contact is nearly open and assume spin is conserved. It will prove useful to follow Ref. 16 and write Eq. (2.13) as a $0+1$ dimensional Euclidean path integral for $\theta_{\rho,\sigma}(\tau) \equiv \theta_{\rho,\sigma}(x=0, \tau)$. This formulation is not essential for carrying out the perturbative analysis of this fixed point. However, it is of conceptual value for discussing the duality between different phases, which can be understood in terms of instanton processes in which $\theta_{\rho,\sigma}(\tau)$ tunnels between degenerate minima at strong coupling. This is accomplished by setting up the path integral for $\theta_{\sigma\rho}(x, \tau)$ and then integrating out $\theta_{\sigma,\rho}(x, \tau)$ for $x \neq 0$. The resulting theory for $\theta_{\sigma,\rho}(\tau)$ has the form of a *quantum Brownian motion* model[24,25,28–30] described by the Euclidean action
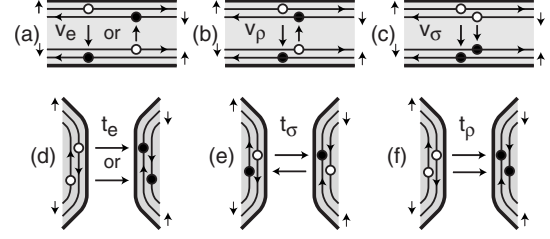
FIG. 2. Schematic representation of tunneling processes in [(a)–(c)] the CC phase (small v) and [(c)–(e)] the II phase (small t). (a) and (d) describe single electron processes, while the others are two particle processes. The duality relating $v_e \leftrightarrow t_e$, $v_\rho \leftrightarrow t_\sigma$, and $v_\sigma \leftrightarrow t_\rho$ can clearly be seen.

$$S_{CC} = \frac{1}{\beta} \sum_{\alpha, \omega_n} \frac{1}{2\pi g_\alpha} |\omega_n| |\theta_a(\omega_n)|^2 - \int_0^\beta \frac{d\tau}{\tau_c} V_{CC}(\theta_\sigma, \theta_\rho), \quad (2.19)$$

where $\omega_n = 2\pi n/\beta$ are Matsubara frequencies and $\beta = 1/k_B T$. We have included the short time cutoff $\tau_c = x_c/v$ in the second term to make the potential $V(\theta_\rho, \theta_\sigma)$ dimensionless. The theory can be regularized by evaluating frequency sums with a $\exp(-|\omega_n|\tau_c)$ convergence factor.

The potential $V(\theta_\rho, \theta_\sigma)$ is given by an expansion in terms of tunneling operators, which represent the processes depicted in Figs. 2(a)–2(c),

$$V_{CC} = v_e \cos(\theta_\rho + \eta_\rho)\cos\theta_\sigma + v_\rho \cos 2\theta_\rho + v_\sigma \cos 2\theta_\sigma. \quad (2.20)$$

v_e represents the elementary backscattering of a single electron across the point contact. The phase of $\cos\theta_\sigma$ in that term is fixed by time-reversal symmetry. The phase η_ρ of $\cos\theta_\rho$ is arbitrary, though mirror symmetry, if present, requires $\eta_\rho = n\pi$. In addition we include compound tunneling processes. v_ρ represents the backscattering of a pair of electrons with opposite spins. We have chosen to define θ_ρ such that the phase of this term is zero. Note that this process involves the tunneling of *spin* (not charge) between the top and bottom edges. Similarly, v_σ represents the transfer of a unit of spin from the right to the left moving channels and involves the tunneling of *charge* $2e$ between the top and bottom edges. In general higher order terms could also be included. However, those terms are less relevant.

The low energy stability of this fixed point is determined by the scaling dimensions $\Delta(v_\alpha)$ of the perturbations, which determine the leading order renormalization group flows,

$$dv_\alpha/d\ell = [1 - \Delta(v_\alpha)]v_\alpha. \quad (2.21)$$

These are given by

$$\Delta(v_e) = (g_\rho + g_\sigma)/2 = (g + g^{-1})/2,$$

$$\Delta(v_\rho) = 2g_\rho = 2g,$$

$$\Delta(v_\sigma) = 2g_\sigma = 2g^{-1}. \quad (2.22)$$

It is therefore clear that all operators are irrelevant for $1/2 < g < 2$, so that the CC phase is stable. For $g < 1/2$ v_ρ be-

comes relevant and for $g > 2$ v_σ becomes relevant.

At the fixed point the conductance matrix elements are

$$G_{XX} = 2e^2/h,$$

$$G_{YY} = G_{XY} = 0. \qquad (2.23)$$

At finite temperature, there will be corrections to these values. The leading corrections will depend on the least irrelevant operators. We find

$$\delta G_{XX} = \begin{cases} -c_1 v_e^2 T^{g+g^{-1}-2}, & g > 1/\sqrt{3} \\ -c_2 v_\rho^2 T^{4g-2}, & g < 1/\sqrt{3}, \end{cases}$$

$$\delta G_{YY} = \begin{cases} c_3 v_e^2 T^{g+g^{-1}-2}, & g < \sqrt{3} \\ c_4 v_\sigma^2 T^{4/g-2}, & g > \sqrt{3}, \end{cases} \qquad (2.24)$$

where c_i are nonuniversal constants. Note that for $g < 1/\sqrt{3}$ the exponents for G_{XX} and G_{YY} are different. In addition, there will be power-law corrections to G_{XY} when the mirror symmetries $\mathcal{M}_x, \mathcal{M}_y$ are violated. However, this correction is zero when computed from Eqs. (2.19) and (2.20) even when $\eta_\rho \neq 0$ due to the symmetry of Eq. (2.20) under $\theta_\sigma \rightarrow -\theta_\sigma$. Computing G_{XY} requires a higher order irrelevant operator. For instance, $\lambda_1 \partial_x \varphi_\sigma \sin \theta_\rho \cos \theta_\sigma$ and $\lambda_2 \partial_x \varphi_\rho \cos \theta_\rho \sin \theta_\sigma$ break both $\mathcal{M}_X$ and $\mathcal{M}_Y$ while preserving time reversal. This leads to

$$\delta G_{XY} = c_5 \lambda_1 \lambda_2 T^{g+g^{-1}}. \qquad (2.25)$$

Note that the temperature exponent of G_{XY} is at least 2 even for weak interactions $g \sim 1$. This is because the tunneling terms λ_1 and λ_2 include an extra derivative term. This is related to the fact (which we will show in Sec. III C) that for noninteracting electrons $G_{XY}=0$. Weak interactions then introduce inelastic processes which give $G_{XY} \propto T^2$. The vanishing of G_{XY} is a unique property of the spin filtered edge states of the QSHI, which does not occur for a generic four-terminal conductance.

2. Weak tunneling (II) limit

When the point contact is pinched off, $\theta_{\rho,\sigma}$ are effectively pinned, and a theory can be developed in terms of electron tunneling process across the point contact. This theory is most conveniently expressed in terms of the *discontinuity* $\tilde{\theta}_{\sigma,\rho} \equiv \varphi_{\sigma,\rho}^{\text{right}} - \varphi_{\sigma,\rho}^{\text{left}}$ across the junction.[31] The theory takes the form

$$S_{\text{II}} = \frac{1}{\beta} \sum_{\alpha,\omega_n} \frac{g_\alpha}{2\pi} |\omega_n| |\tilde{\theta}_a(\omega_n)|^2 - \int_0^\beta \frac{d\tau}{\tau_c} V_{\text{II}}(\tilde{\theta}_\sigma, \tilde{\theta}_\rho),$$

$$(2.26)$$

with

$$V_{\text{II}} = t_e \cos(\tilde{\theta}_\rho + \eta_\rho) \cos \tilde{\theta}_\sigma + t_\rho \cos 2\tilde{\theta}_\rho + t_\sigma \cos 2\tilde{\theta}_\sigma. \qquad (2.27)$$

As depicted in Figs. 2(d)–2(f) t_e represents the tunneling of a single electron from left to right across the junction. t_σ de-

scribes the transfer of a unit of spin across the junction. t_ρ describes the tunneling of a pair of electrons with opposite spins.

The relationship between S_{II} and S_{CC} can be understood into two ways. First, since both S_{II} and S_{CC} describe tunneling between the middles of two disconnected Luttinger liquids (either on the top and bottom of the junction or the left and right) the two theories are identical. It is straightforward to see that if we make the identification

$$\theta_\rho \leftrightarrow \tilde{\theta}_\sigma,$$

$$\theta_\sigma \leftrightarrow \tilde{\theta}_\rho, \qquad (2.28)$$

it follows that

$$S_{\text{II}}(g_\rho, g_\sigma, t_e, t_\rho, t_\sigma) = S_{\text{CC}}(g_\sigma, g_\rho, v_e, v_\sigma, v_\rho). \qquad (2.29)$$

Thus, the small v and small t theories are dual to each other, with the identification

$$v_e \leftrightarrow t_e,$$

$$v_\rho \leftrightarrow t_\sigma,$$

$$v_\sigma \leftrightarrow t_\rho,$$

$$g \leftrightarrow g^{-1}. \qquad (2.30)$$

Using this identification, the scaling dimensions $\Delta(t_\alpha)$ can be read off from Eq. (2.22). Thus, like the CC phase, the II phase is stable when $1/2 < g < 2$. The low temperature conductance can also be read from Eqs. (2.23)–(2.25) using the identification

$$G_{XX} \leftrightarrow G_{YY}. \qquad (2.31)$$

Another way to understand this duality, which will prove useful below, is to consider an instanton expansion for strong coupling. For large v_e $(\theta_\rho, \theta_\sigma)$ will be tightly bound at the minima of $V(\theta_\rho, \theta_\sigma)$, shown in Fig. 3(a). (Here we assume for simplicity $\eta_\rho = 0$.) The partition function describing the path integral of Eq. (2.19) can then be expanded in instanton processes, in which $(\theta_\rho, \theta_\sigma)$ switches between nearby minima at discrete times. Evaluating the first term in Eq. (2.19) for a configuration of instantons leads to an interaction between the instantons which depends logarithmically on time. The expansion describes the partition function for a one-dimensional "Coulomb gas," where the "charges" correspond to the tunneling events. This Coulomb gas has exactly the same form as the expansion of Eq. (2.26) in powers of t_e, t_ρ, and t_σ. Thus, we can identify t_e, t_ρ, and t_σ as the fugacities of the instantons.

This duality argument also works in reverse. Starting from Eq. (2.26) we can derive Eq. (2.19) by considering large t_e and expanding in instantons in $\tilde{\theta}_\rho$ and $\tilde{\theta}_\sigma$ connecting minima in Fig. 3(b), which have fugacities v_e, v_ρ, and v_σ.

3. Charge insulator/spin conductor (IC)

We next study the mixed charge insulator spin conductor phase. To generate the effective action for this phase, includ-
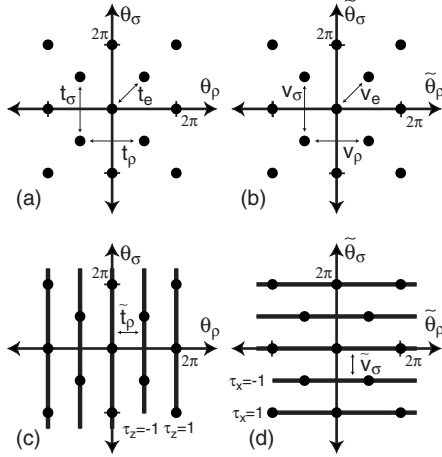
FIG. 3. (a) Positions of the minima of $V(\theta_\rho, \theta_\sigma)$ in Eq. (2.20). When the minima are deep instanton tunneling events between the minima, denoted by t_e, t_ρ, and t_σ, correspond to the transfer of charge and spin across the junction and define the dual theory [Eqs. (2.26) and (2.27)]. (b) Positions of the minima of $V(\widetilde{\theta}_\rho, \widetilde{\theta}_\sigma)$ in the dual theory [Eqs. (2.26) and (2.27)]. Instanton process v_e, v_ρ, and v_σ correspond to backscattering of charge and spin in the original theory. (c) The IC phase viewed from the CC limit. When v_ρ is large and $v_\sigma=0$, the minima of $V(\theta_\rho, \theta_\sigma)$ in Eq. (2.20) are on one-dimensional valleys and define the IC phase. When v_σ is small but finite the valleys have a periodic potential $\widetilde{v}_\sigma \tau^z \cos \theta_\sigma$, with opposite signs $\tau^z = \pm 1$ in neighboring valleys. Instanton tunneling processes between the valleys, denoted $\widetilde{t}_\rho$, switch the sign of τ^z. (d) The IC phase viewed from the II limit, in which $t_\rho=0$ and t_σ is large. The valleys have periodic potential $\widetilde{t}_\rho \tau^x \cos \widetilde{\theta}_\rho$ with $\tau^x=\pm 1$, whose sign is switched by instanton processes $\widetilde{v}_\sigma$.

ing the leading relevant operators, it is useful to use the instanton analysis discussed at the end of Sec. II B 2. Consider Eqs. (2.26) and (2.27) for large v_ρ, keeping v_e and v_σ small. θ_ρ will be pinned in the minima of $-\cos 2\theta_\rho$, $\theta_\rho=n\pi$, while θ_σ remains free to fluctuate. $(\theta_\rho, \theta_\sigma)$ are thus confined to "valleys" along the vertical lines in Fig. 3(c).

There are two types of perturbations to be considered. First, v_e will lead to a periodic potential along the vertical lines, with minima at the dots. Note, however, that on alternate lines the sign of the periodic potential changes since $\cos \theta_\rho \cos \theta_\sigma \sim (-1)^n \cos \theta_\sigma$ for $\theta_\rho=n\pi$.

Next consider an instanton process where θ_ρ tunnels between neighboring valleys. In this process, $\theta_\rho \to \theta_\rho \pm \pi$, but θ_σ is unchanged. It follows that the v_e perturbation discussed above changes sign. Thus, the instanton process does not commute with the v_e term.

The expansion of the partition function in both instantons and v_e can be generated by the action for the IC phase given by $S_{IC}=S_{IC}^0+S_{IC}^1$ with

$$S_{IC}^0 = \frac{1}{\beta}\sum_{\omega_n} \frac{g_\rho}{2\pi}|\omega_n||\widetilde{\theta}_\rho(\omega_n)|^2 + \frac{1}{2\pi g_\sigma}|\omega_n||\theta_\sigma(\omega_n)|^2$$

(2.32)

and

$$S_{IC}^1 = \int_0^\beta \frac{d\tau}{\tau_c}[\widetilde{t}_\rho \tau^x \cos \widetilde{\theta}_\rho + \widetilde{v}_\sigma \tau^z \cos \theta_\sigma].$$ (2.33)

Here $\widetilde{t}_\rho$ describes the instanton tunneling process. The tilde distinguishes it from the ordinary charge tunneling process, which involves charge $2e$. $\widetilde{t}_\rho$ describes a tunneling of charge e *without* spin. $\widetilde{v}_\sigma$ describes the periodic potential as a function of θ_σ generated by v_e. We have introduced a pseudospin degree of freedom $\tau^z = \pm 1$ to account for the sign of $\cos \theta_\rho$ in the different valleys. Since the instanton process switches the sign, it is associated with τ^x. Expanding the partition function defined by Eqs. (2.32) and (2.33) in powers of $\widetilde{t}_\rho$ and $\widetilde{v}_\sigma$ precisely generates the expansion of Eqs. (2.19) and (2.20) in instantons.

It is also instructive to derive Eqs. (2.32) and (2.33) starting from the opposite limit of the II phase described by Eqs. (2.26) and (2.27). In this case, consider large t_σ, which leads to the horizontal valleys as a function of $\widetilde{\theta}_\rho$ and $\widetilde{\theta}_\sigma$ in Fig. 3(d). The roles of the two terms in Eq. (2.33) are thus reversed. $\widetilde{t}_\rho$ describes the periodic potential along the valleys, which has a sign specified by $\tau^x = \pm 1$. $\widetilde{v}_\sigma$ describes the instanton processes which switch the sign of τ^x.

The lowest order renormalization group flows depend only on the scaling dimensions of $\widetilde{t}_\rho$ and $\widetilde{v}_\sigma$ and are unaffected by the pseudospin $\tau^{x,z}$. We find

$$\Delta(\widetilde{t}_\rho) = \frac{1}{2g_\rho} = \frac{1}{2g},$$

$$\Delta(\widetilde{v}_\sigma) = \frac{g_\sigma}{2} = \frac{1}{2g}.$$ (2.34)

Thus, the IC phase is stable when $g<1/2$.

In Sec. III D we will require the renormalization group flow to third order in $\widetilde{t}_\rho$ and $\widetilde{v}_\sigma$. There, the nontrivial interaction between them introduced by the pseudospin will play a crucial role.

The conductivity at the IC fixed point is given by

$$G_{XX} = G_{YY} = G_{XY} = 0.$$ (2.35)

This, however, does *not* mean that the full four-terminal conductance is zero because spin conservation still requires $G_{ZZ}=2e^2/h$. This leads to the nontrivial structure in the four-terminal conductance predicted in Ref. 10.

At finite temperature, there will be corrections to the conductance. We find

$$\delta G_{XX} = d_1 \widetilde{t}_\rho^2 T^{g^{-1}-2},$$

$$\delta G_{YY} = d_2 \widetilde{v}_\sigma^2 T^{g^{-1}-2}.$$ (2.36)

As in this section the corrections to G_{XY} will depend on a higher order irrelevant operator. For instance, $\lambda_1 \tau^y \sin \widetilde{\theta}_\rho \sin \theta_\sigma$ and $\lambda_2 \tau^y \cos \widetilde{\theta}_\rho \cos \theta_\sigma$ lead to

$$\delta G_{XY} = d_3 \lambda_1 \lambda_2 T^{2g^{-1}-2}. \tag{2.37}$$

As in Eq. (2.25), G_{XY} is suppressed more strongly at low temperature than G_{XX} and G_{YY}, and the exponent is larger than 2 for $g < 1/2$.

4. Charge conductor/spin insulator (CI)

For $g > 2$ the perturbation $v_\sigma \cos 2\theta_\sigma$ in Eq. (2.20) becomes relevant and drives the system to the CI phase. This may be described in a manner similar to the IC phase. It is described by the action $S_{CI} = S_{CI}^0 + S_{CI}^1$ with

$$S_{CI}^0 = \frac{1}{\beta} \sum_{\omega_n} \frac{g_\sigma}{2\pi} |\omega_n| |\widetilde{\theta}_\sigma(\omega_n)|^2 + \frac{1}{2\pi g_\rho} |\omega_n| |\theta_\rho(\omega_n)|^2 \tag{2.38}$$

and

$$S_{CI}^1 = \int_0^\beta \frac{d\tau}{\tau_c} [\widetilde{t}_\sigma \tau^z \cos(\widetilde{\theta}_\sigma + \eta_\sigma) + \widetilde{v}_\rho \tau^x \cos(\theta_\rho + \eta_\rho)]. \tag{2.39}$$

The leading relevant operators have dimensions

$$\Delta(\widetilde{t}_\sigma) = \frac{1}{2g_\sigma} = \frac{g}{2},$$

$$\Delta(\widetilde{v}_\rho) = \frac{g_\rho}{2} = \frac{g}{2}. \tag{2.40}$$

This phase is thus stable when $g > 2$ and has conductances

$$G_{XX} = G_{YY} = 2e^2/h,$$

$$G_{XY} = 0. \tag{2.41}$$

5. Spin-orbit interactions and T-breaking insulator

In this section we consider the role of spin-orbit interaction terms which violate the conservation of spin S_z but respect time-reversal symmetry. We will argue that such terms are irrelevant for the critical behavior of the point contact when $g > 1/2$, but they are relevant for $g < 1/2$ and drive the system at low energy to a TBI.

Time-reversal symmetry allows the following terms in the expansion about CC fixed point (2.19):

$$S_{CC}^{SO} = \int_0^\beta \frac{d\tau}{\tau_c} [v_{so} \cos \varphi_\sigma \sin \theta_\sigma + v_{sf} \cos(2\varphi_\sigma + \eta_{sf})]. \tag{2.42}$$

The first term is a single electron process $\psi_{R\uparrow}^\dagger \psi_{R\downarrow}$ [Fig. 4(a)] in which an electron flips its spin and crosses the junction. The second term is a correlated tunneling process $\psi_{R\uparrow}^\dagger \psi_{L\uparrow}^\dagger \psi_{R\downarrow} \psi_{L\downarrow}$ [Fig. 4(b)], where a left and right moving pair of up spins flip into a left and right moving pair of down spins. Referring to Table I, it is clear that both terms respect time-reversal symmetry. η_{sf} is allowed by time-reversal symmetry but violates both mirrors $\mathcal{M}_x$ and $\mathcal{M}_y$. Higher order
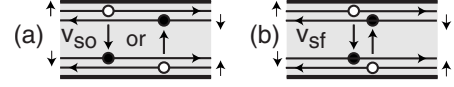
FIG. 4. Tunneling processes in the CC limit allowed by spin nonconserving spin-orbit interactions. v_{so} is a single particle process where a single spin is flipped, while v_{sf} is a two particle process, flipping two spins.

processes are also possible, though they will be less relevant perturbatively.

It is straightforward to determine the scaling dimensions of these perturbations. We find

$$\Delta(v_{so}) = \frac{1}{2}(g_\sigma + g_\sigma^{-1}) = \frac{1}{2}(g + g^{-1}),$$

$$\Delta(v_{sf}) = \frac{2}{g_\sigma} = 2g. \tag{2.43}$$

For $g \neq 1$ the single particle spin-orbit term v_{so} is always irrelevant. However, v_{sf} becomes relevant when $g < 1/2$.

At finite temperature these lead to corrections to the conductance of the CC phase. To lowest order they do not affect G_{XX}, G_{XY}, and G_{YY}. However we find

$$\delta G_{ZZ} \propto \begin{cases} T^{g+g^{-1}-2}, & g > 1/\sqrt{3} \\ T^{4g-2}, & g < 1/\sqrt{3}. \end{cases} \tag{2.44}$$

Like G_{XY}, G_{ZX} and G_{ZY} are zero unless higher order irrelevant operators, which involve extra powers of $\partial_x \varphi_\alpha$ or $\partial_x \theta_\alpha$, are included. We find

$$\delta G_{ZX} \propto T^{2g},$$

$$\delta G_{ZY} \propto T^{g+g^{-1}}. \tag{2.45}$$

For weak interactions, $g \sim 1$, these conductances vanish for $T \to 0$ as T^2.

For $g < 1/2$ there are two relevant perturbations about the CC limit. To study their effects we consider a model in which only the relevant perturbations appear. Since these perturbations involve the commuting operators φ_σ and θ_ρ, it is useful to study the $0+1$ dimensional field theory of those variables,

$$S_{CC}^0 = \sum_{\omega_n} \frac{1}{2\pi g_\rho} |\omega_n| |\theta_\rho(\omega_n)|^2 + \frac{g_\sigma}{2\pi} |\omega_n| |\varphi_\sigma(\omega_n)|^2, \tag{2.46}$$

with

$$S_{CC}^1 = \int_0^\beta \frac{d\tau}{\tau_c} [v_\rho \cos(2\theta_\rho + \eta_\rho) + v_{sf} \cos(2\varphi_\sigma + \eta_{sf})]. \tag{2.47}$$

The low temperature behavior of this theory can be studied by the duality arguments of Sec. II B 2. When v_ρ and v_{sf} are both large, $(\theta_\rho, \varphi_\sigma)$ will be stuck in the deep minima of $V_{CC}(\theta_\rho, \varphi_\sigma)$ shown in Fig. 5. In this phase, the *four-terminal* conductance is zero,
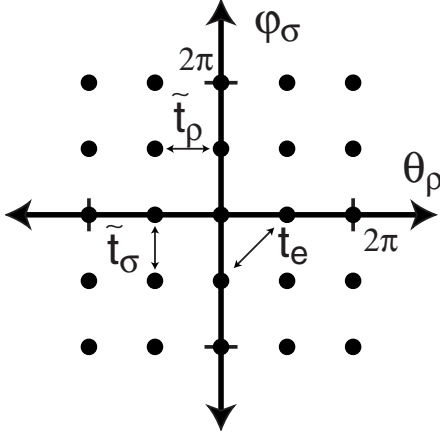
FIG. 5. Minima of the potential $V(\theta_\rho, \varphi_\sigma)$ in Eq. (2.47). Large v_ρ and v_{so} define the time-reversal breaking insulating phase. Instanton processes $\tilde{t}_\rho$ and $\tilde{t}_\sigma$ restore time-reversal invariance. They correspond to tunneling of spinless chargons or chargeless spinons.

$$G_{AB} = 0. \tag{2.48}$$

This can be seen most simply by renaming the variables

$$\theta_\rho \to \theta_1 + \theta_2,$$

$$\varphi_\sigma \to \theta_1 - \theta_2,$$

$$\varphi_\rho \to \varphi_1 + \varphi_2,$$

$$\varphi_\sigma \to \varphi_1 - \varphi_2. \tag{2.49}$$

The interpretation of $\theta_{1(2)}$ and $\varphi_{1(2)}$ is simple. They are the usual Luttinger liquid charge and phase variables for the top (bottom) edges in Figs. 2(a)–2(c). In the strong coupling phase θ_1 and θ_2 are both pinned, so that any current flowing in from any lead is perfectly reflected back into that lead. The four leads are completely decoupled.

This is the same perfectly reflecting phase that would arise if we had a single particle backscattering term on each edge $v_{back}(\cos 2\theta_1 + \cos 2\theta_2) = 2v_{back} \cos \theta_\rho \cos \varphi_\sigma$, which would be relevant for $g < 1$. However in our problem that term is forbidden by time-reversal symmetry. It is thus clear that time-reversal symmetry is violated by the strong coupling fixed point. It is useful to see this from Fig. 5. Note that since under time reversal $\varphi_\sigma \to \varphi_\sigma + \pi$. Thus pinning φ_σ violates time reversal. There are two sets of minima of $V(\theta_\rho, \varphi_\sigma)$ which are interchanged by the time-reversal operation.

At finite temperature tunneling processes between the two sets of minima of $V(\theta_\rho, \varphi_\sigma)$ will restore time-reversal symmetry. These instanton processes correspond to tunneling of charge from one lead to another. Interestingly, the lowest order instanton processes, denoted $\tilde{t}_\rho$ and $\tilde{t}_\sigma$, do not correspond to tunneling of electrons but rather spinless charge e "chargons" or charge neutral "spinons."

The scaling dimensions of these instanton processes can be deduced from Eqs. (2.46) and (2.47). We find

$$\Delta(\tilde{t}_\rho) = \frac{1}{2g_\rho} = \frac{1}{2g},$$

$$\Delta(\tilde{t}_\sigma) = \frac{g_\sigma}{2} = \frac{1}{2g}. \tag{2.50}$$

Thus, both processes are irrelevant for $g < 1/2$, and the TBI phase is stable. These processes lead to power-law temperature behavior,

$$\delta G_{XX} = c_1 \tilde{t}_\rho^2 T^{1/g-2},$$

$$\delta G_{YY} = c_1 \tilde{t}_\sigma^2 T^{1/g-2}. \tag{2.51}$$

When the $\tilde{t}_{\rho,\sigma}$ processes dominate, there will be nontrivial noise correlations in the current. The $\tilde{t}_\rho$ process involves transferring charge $e/2$ from lead 1 to lead 2 and another $e/2$ from lead 4 to lead 3. This leads to correlations in the low frequency noise defined by

$$\mathcal{S}_{ij}(\omega) = \int dt e^{i\omega t} \langle I_i(t)I_j(0) + I_j(0)I_i(t) \rangle. \tag{2.52}$$

Consider the two-terminal geometry $I_X = G_{XX}V_X$. The current I_X will be carried by the $\tilde{t}_\rho$ processes, so that $I_1 = I_4 = I_X/2$. The shot noise correlations in the limit $\omega \to 0$ will be

$$S_{11} = S_{44} = S_{14} = S_{41} = 2e^*I_1 \tag{2.53}$$

with $e^* = e/2$. Thus, the currents are all perfectly correlated, and the current in each lead is carried by *fractional* charges, $e/2$.

III. CRITICAL BEHAVIOR OF CONDUCTANCE

In this section we describe the critical behavior of the conductance at the pinch-off transition of the point contact. We will compute the critical conductance G_g^*, the critical exponent α_g, and the scaling function $\mathcal{G}_g(X)$ in certain solvable limits. We will begin in Sec. III A with a discussion of the general properties of the scaling function and a summary of our calculated results. Then in the following sections we will describe in detail our calculations for $g = 1 - \epsilon$, $g = 1/\sqrt{3}$, and $g = 1/2 + \epsilon$.

A. Scaling behavior and summary of results

The stability analysis of the previous sections leads to the phase diagram as a function of g depicted in Fig. 6(a). The top line depicts the CC phase and the bottom line depicts the II phase, and the arrows denote the stability associated with the leading relevant operators. Since the II and CC phases are both stable for $1/2 < g < 2$ they are separated by an intermediate unstable fixed point P, denoted by the dashed central line. For $g < 1/2$ the II and CC phases become unstable, and when spin is conserved the flow is toward the IC phase. We will see in Sec. III D that the unstable critical fixed point matches smoothly onto the IC fixed point at $g = 1/2$. Similarly, the CI fixed point is stable for $g > 2$ and connects to the critical fixed point at $g = 2$.

For $1/2 < g < 2$ the unstable intermediate fixed point P describes the critical behavior of the pinch-off transition of the point contact. We will argue that this fixed point is characterized by a *single* relevant operator, which allows us to
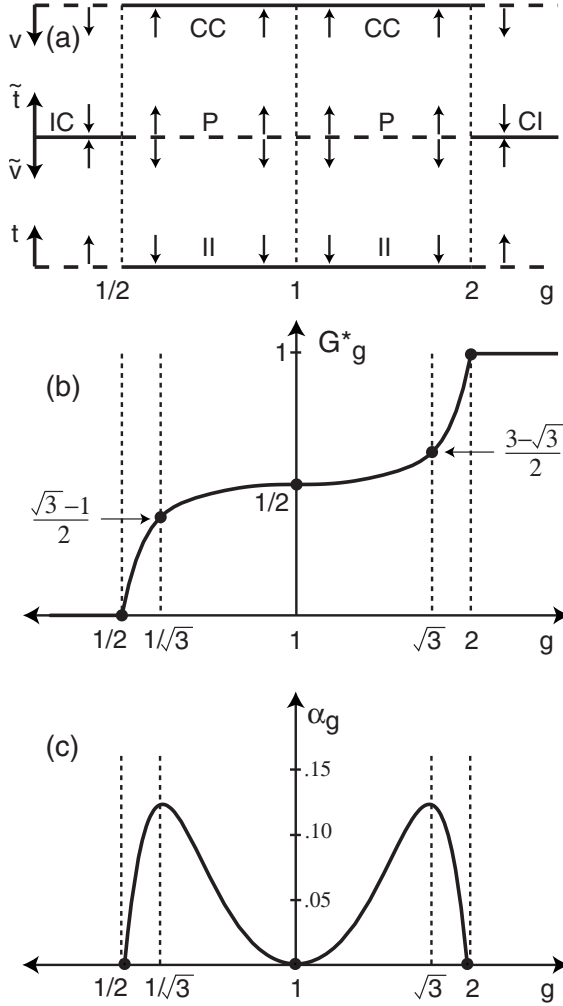
FIG. 6. (a) Phase diagram for a point contact in a QSHI as a function of the Luttinger liquid parameter g. The arrows indicate the stability of the CC, II, CI, and IC phases, as well as the critical fixed point P. This figure assumes spin conservation. In the presence of spin-orbit interactions, the IC phase is unstable for $g < 1/2$. This leads to the TBI phase discussed in Sec. II B 5. (b) Conductance G^* of the critical fixed point as a function of g. The curve is a fit, which incorporates the data in Eq. (3.7). (c) Critical exponent α_g as a function of g. The curve is a fit incorporating the data in Eq. (3.8). g is plotted on a logarithmic scale in all three panels to emphasize the $g \leftrightarrow 1/g$ symmetry.

formulate a single parameter scaling theory for the pinch-off transition. If we denote u as the relevant operator, then the leading order renormalization group flow near the fixed point has the form

$$du/d\ell = \alpha_g u, \qquad (3.1)$$

where α_g is a critical exponent to be determined. By varying a gate voltage V_G it is possible to cross from the region of stability of the II phase to the region of stability of the CC phase. In the process one must pass through the fixed point $u=0$ at $V_G=V_G^*$. Near the transition, we thus have $u \propto \Delta V_G = V_G - V_G^*$. Under a renormalization group transformation in which energy, length and time are rescaled by b, we have

$u \to u b^{\alpha_g}$ and $T \to Tb$. Invariance under this transformation requires that physical quantities can only depend on u and T in the combination u/T^{α_g}. Close to the transition we thus have

$$\lim_{T, \Delta V_G \to 0} G_{AB}(T, \Delta V_G) = 2\frac{e^2}{h}\mathcal{G}_{g,AB}\left(c\frac{\Delta V_G}{T^{\alpha_g}}\right), \qquad (3.2)$$

where c is a nonuniversal constant and $\mathcal{G}_{g,AB}$ is a universal crossover scaling function which varies between 0 and 1.

We will argue that the critical point characterizing the pinch-off transition has emergent spin conservation as well as mirror symmetry, so that the only nonzero elements of the conductance matrix are G_{XX} and G_{YY}. Moreover, the duality considerations discussed in Sec. III C require that $\mathcal{G}_{g,YY}(X)$ and $\mathcal{G}_{g,XX}(X)$ are related, so that they are both determined by the same universal scaling function,

$$\mathcal{G}_{g,XX}(X) = \mathcal{G}_g(X),$$

$$\mathcal{G}_{g,YY}(X) = \mathcal{G}_g(-X). \qquad (3.3)$$

The scaling function $\mathcal{G}_g(X)$ has some general properties which are easy to deduce. First, the equivalence between the CC theory at g with the II theory at $1/g$ leads to the relation

$$\mathcal{G}_{1/g}(X) = 1 - \mathcal{G}_g(-X). \qquad (3.4)$$

Second, when $T \to 0$ for fixed ΔV_G the system flows to either the CC or the II phase, where the temperature dependence of the conductance is given by Eq. (2.24). The behavior of the scaling function for large X then follows

$$\mathcal{G}_g(X \to +\infty) = 1 - a_g^+ X^{-\beta_g^+},$$

$$\mathcal{G}_g(X \to -\infty) = a_g^- X^{-\beta_g^-}. \qquad (3.5)$$

The coefficients $a_g^{\pm}$ depend on the normalization of X but can be fixed if we specify $\mathcal{G}_g'(X=0)=1/2$. The exponents obey the relations

$$\beta_g^+ = \begin{cases} (4g-2)/\alpha_g, & 1/2 < g < 1/\sqrt{3} \\ (g+g^{-1}-2)/\alpha_g, & 1/\sqrt{3} < g < 1, \end{cases}$$

$$\beta_g^- = (g+g^{-1}-2)/\alpha_g \quad 1/2 < g < 1. \qquad (3.6)$$

The behavior of $\beta_g^{\pm}$ for $1 < g < 2$ can be deduced using Eq. (3.4).

In the following sections we compute properties of the scaling function at $g=1-\epsilon$, $g=1/\sqrt{3}$, and $g=1/2+\epsilon$. From Eq. (3.4) we can deduce corresponding results at $g=1+\epsilon$, $g=\sqrt{3}$, and $g=2-\epsilon$. First consider the critical conductance $G_g^*=\mathcal{G}_g(X=0)$. We find

$$G_g^* = \begin{cases} 1/2 + O(\epsilon^3), & g=1-\epsilon \\ (\sqrt{3}-1)/2, & g=1/\sqrt{3} \\ \pi^2\epsilon, & g=1/2+\epsilon. \end{cases} \qquad (3.7)$$

These results are summarized in Fig. 6(b). The curve is a polynomial fit of $G^*(\ln g)$ which incorporates the data in Eq. (3.7) and the $g \leftrightarrow 1/g$ symmetry. It is satisfying that the
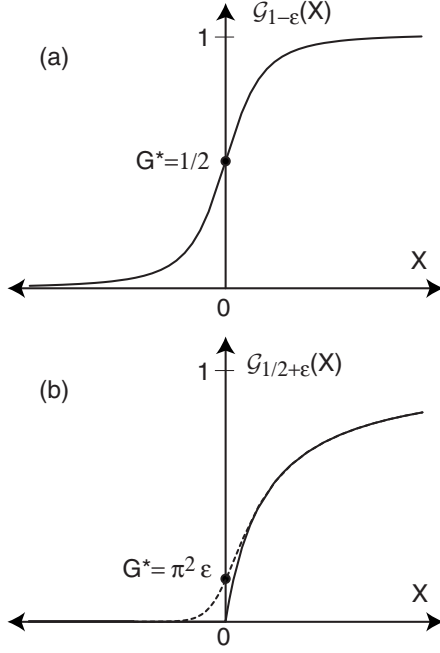
FIG. 7. The universal scaling function $\mathcal{G}_g(X)$ for (a) $g = 1 - \epsilon$ [Eq. (3.9)] and (b) $g = 1/2 + \epsilon$ [Eq. (3.10)]. In (b) the solid line is $\epsilon \to 0$, and the dashed line shows the approximate behavior for $\epsilon \sim .02$.

curve is smooth and monotonic, which indicates a consistency between the slopes at $g = 1/2, 1$ and the value at $g = 1/\sqrt{3}$.

We are able to deduce the critical exponent α_g for $g = 1 - \epsilon$ and $g = 1/2 + \epsilon$. We find

$$\alpha_g = \begin{cases} \epsilon^2/2, & g = 1 - \epsilon \\ 4\epsilon, & g = 1/2 + \epsilon. \end{cases} \quad (3.8)$$

These results are summarized in Fig. 6(c). The curve is a polynomial fit of $\alpha(\ln g)$. It is suggestive that in this fit α_g exhibits a maximum near $g = 1/\sqrt{3}$ with a value $\alpha_{1/\sqrt{3}} = 0.123 \sim 1/8$. It is possible, however, that α_g exhibits a cusp at $g = 1/\sqrt{3}$ analogous to the behavior of β_g in Eq. (3.6).

In Secs. III C and III D we compute the full scaling function $\mathcal{G}_g(X)$ in the limits $g = 1 - \epsilon$ and $g = 1/2 + \epsilon$ to lowest order in ϵ. For $g = 1 + \epsilon$, $\epsilon \to 0$ we find

$$\mathcal{G}_1(X) = \frac{1}{2}\left(1 + \frac{X}{\sqrt{1 + X^2}}\right). \quad (3.9)$$

For $g = 1/2 + \epsilon$, $\epsilon \to 0$,

$$\mathcal{G}_{1/2}(X) = \theta(X)\frac{X}{1 + X}. \quad (3.10)$$

The singular behavior near $X = 0$ in Eq. (3.10) is rounded for finite ϵ. The perturbative analysis in Sec. III D 1 shows that for $|X| \ll 1$

$$\mathcal{G}_{1/2 + \epsilon}(X) = \frac{X}{1 - e^{-X/(\pi^2 \epsilon)}}. \quad (3.11)$$

$\mathcal{G}_{1-\epsilon}(X)$ and $\mathcal{G}_{1/2+\epsilon}(X)$ are plotted in Figs. 7(a) and 7(b). For g close to 1 the pinch-off curve is symmetrical about G^*

TABLE II. Parameters in Eq. (3.5) for the asymptotic behavior of the scaling function $\mathcal{G}_g(X)$ in the solvable limits $g \to 1$, $g \to 1/2$.

g	β_g^+	a_g^+	β_g^-	a_g^-
$1 - \epsilon$	2	1/4	2	1/4
$1/2 + \epsilon$	1	1	$1/(8\epsilon)$	$(2.75)^{1/(8\epsilon)}$

$= e^2/h$. However, for stronger repulsive interactions it becomes asymmetrical, as G^* is reduced, approaching 0 at $g = 1/2$.

The asymptotic $|X| \to \infty$ behavior (3.5) of $\mathcal{G}_1(X)$ and $\mathcal{G}_{1/2+\epsilon}$ can also be determined from Eqs. (3.9) and (3.10) though a separate calculation (see Sec. III D 3) is required for $\mathcal{G}_{1/2+\epsilon}(X \to -\infty)$. The results, which are consistent with Eq. (3.6), are shown in Table II.

B. Quantum Brownian motion model, duality, and $g = 1/\sqrt{3}$

In this section we recast the Luttinger liquid model as a model of QBM in a periodic potential. This mapping elucidates the duality between the CC and II limits and exposes an extra symmetry, the problem at $g = 1/\sqrt{3}$, which allows us to deduce the critical conductance at that point. We begin with a brief review of the QBM model and then derive its consequences for the scaling functions $\mathcal{G}_g(X)$ and $G^*_{1/\sqrt{3}}$.

1. Quantum Brownian motion model

The QBM model[28–30] was originally formulated as a theory of the motion of a heavy particle coupled to an Ohmic dissipative environment modeled as a set of Caldeira-Leggett oscillators.[32] Though the applicability of this model to the motion of a real particle coupled to phonons or electron-hole pairs has been questioned,[33,34] it was later shown that this model is directly relevant to quantum impurity problems. Specifically, the QBM model in a one-dimensional periodic potential is equivalent to the theory of a weak link in a single-channel Luttinger liquid.[16,18] In this mapping the QBM takes place in an abstract space where the "coordinate" of the "particle" is the number of electrons that have tunneled past the weak link. The periodic potential is due to the discreteness of the electron's charge. The low energy excitations of the Luttinger liquid play the role of the dissipative bath, and the strength of the dissipation is related to the Luttinger liquid parameter g. The one-dimensional QBM model has two phases: a localized phase with conductance $G = 0$ stable for $g < 1$ and a fully coherent phase with perfect conductance stable for $g > 1$.

The SLL model corresponds to a QBM model in a two-dimensional periodic potential, where the coordinates are the spin and charge variables $\theta_{\rho,\sigma}$. This model is richer than its one-dimensional counterpart because it admits additional fixed points which are intermediate between localized and perfect. These fixed points were first found in the Luttinger liquid model[15,16] and later formulated in terms of the QBM.[25] For certain values of g_ρ and g_σ these intermediate fixed points are related to the three-channel Kondo problem[25] and the three state Potts models.[24] However, those limits are not

directly applicable to the QSHI model, where $g_\rho = 1/g_\sigma = g$. We will show that when $g = 1/\sqrt{3}$ the critical fixed point of the QSHI point contact corresponds to the intermediate point discussed in Ref. 25 for a QBM model on a triangular lattice.

To formulate the QBM model we begin with action (2.19) and define new rescaled variables,

$$\theta_\alpha = \pi\sqrt{2g_\alpha}\,r_\alpha. \tag{3.12}$$

Then Eq. (2.19) takes the form

$$S = \frac{1}{4\pi\beta}\sum_n |\omega_n||\mathbf{r}(\omega_n)|^2 - \int\frac{d\tau}{\tau_c}\sum_\mathbf{G} v_\mathbf{G} e^{2\pi i\mathbf{G}\cdot\mathbf{r}(\tau)}. \tag{3.13}$$

The periodic potential is characterized by reciprocal lattice vectors $\mathbf{G} = m_1\mathbf{b}_1 + m_2\mathbf{b}_2$. The primitive reciprocal lattice vectors $\mathbf{b}_{1,2}$ correspond to the single electron back scattering processes and are given by

$$\mathbf{b}_1 = \frac{1}{\sqrt{2}}(\sqrt{g_\rho}, \sqrt{g_\sigma}), \quad \mathbf{b}_2 = \frac{1}{\sqrt{2}}(\sqrt{g_\rho}, -\sqrt{g_\sigma}). \tag{3.14}$$

The Fourier components of the periodic potential are $v_{\mathbf{b}_1} = v_{\mathbf{b}_2} = v_e e^{i\eta_\rho}/4$, $v_{\mathbf{b}_1+\mathbf{b}_2} = v_\rho/2$, and $v_{\mathbf{b}_1-\mathbf{b}_2} = v_\sigma/2$.

The dual theory is obtained by expanding the partition function for large $v_\mathbf{G}$ in powers of instantons. When $v_\mathbf{G}$ is large, the potential has minima on a real space lattice $\mathbf{R} = n_1\mathbf{a}_1 + n_2\mathbf{a}_2$. The primitive lattice vectors satisfy $\mathbf{a}_i\cdot\mathbf{b}_j = \delta_{ij}$ and are given by

$$\mathbf{a}_1 = \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{g_\rho}}, \frac{1}{\sqrt{g_\sigma}}\right), \quad \mathbf{a}_2 = \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{g_\rho}}, -\frac{1}{\sqrt{g_\sigma}}\right). \tag{3.15}$$

The expansion in instantons connecting these minima is generated by the action

$$S = \frac{1}{4\pi\beta}\sum_n |\omega_n||\mathbf{k}(\omega_n)|^2 - \sum_\mathbf{R}\int\frac{d\tau}{\tau_c}t_\mathbf{R} e^{2\pi i\mathbf{R}\cdot\mathbf{k}(\tau)}. \tag{3.16}$$

This is equivalent to Eqs. (2.26) and (2.27) with $k_\alpha = \pi\sqrt{g_\alpha/2}\,\tilde{\theta}_\alpha$ and $t_{\mathbf{a}_1} = t_{\mathbf{a}_2} = t_e e^{i\eta_\rho}/4$, $t_{\mathbf{a}_1+\mathbf{a}_2} = t_\rho/2$, and $t_{\mathbf{a}_1-\mathbf{a}_2} = t_\sigma/2$.

With the above normalizations for $\mathbf{r}$ and $\mathbf{k}$ the scaling dimensions of the potential perturbations are

$$\Delta(v_\mathbf{G}) = |\mathbf{G}|^2, \quad \Delta(t_\mathbf{R}) = |\mathbf{R}|^2. \tag{3.17}$$

Since operators are relevant when $\Delta < 1$, the most relevant potentials are those with the smallest lattice (reciprocal lattice) vectors $|\mathbf{R}_{\min}|$ ($|\mathbf{G}_{\min}|$). As shown in Refs. 16 and 25 there are ranges of g_ρ and g_σ where both $|\mathbf{R}_{\min}|$ and $|\mathbf{G}_{\min}| > 1$, so that both phases are perturbatively stable. An unstable intermediate fixed point must therefore be present between them.

This fixed point can be accessed perturbatively when $|\mathbf{R}_{\min}|$ and $|\mathbf{G}_{\min}|$ are close to 1. While this does not occur in the regime $g_\rho = 1/g_\sigma$ relevant to the QSHI problem, it is instructive to study this perturbation theory because it provides

evidence that the critical fixed point has emergent mirror and spin conservation symmetry.

When $g_\rho = 1/2 + \epsilon_\rho$ and $g_\sigma = 3/2 + \epsilon_\sigma$ the period potential has triangular symmetry, which is slightly distorted if $\epsilon_\sigma \neq 3\epsilon_\rho$. If we denote the relevant variables as $v_1 = v_{\mathbf{b}_1} = v_{\mathbf{b}_2} = v_e e^{i\eta_\rho}/4$ and $v_2 = v_{\mathbf{b}_1+\mathbf{b}_2} = v_\rho/2$, the second-order renormalization group flow equations are[16]

$$dv_1/d\ell = \frac{1}{2}(\epsilon_\rho + \epsilon_\sigma)v_1 - 2v_1^*v_2,$$

$$dv_2/d\ell = 2\epsilon_\rho v_2 - 2v_1^2. \tag{3.18}$$

These equations describe an intermediate fixed point with a single unstable direction at $v_1 = \sqrt{\epsilon_\rho(\epsilon_\rho + \epsilon_\sigma)}/2$ and $v_2 = (\epsilon_\rho + \epsilon_\sigma)/4$. Note that at the critical point v_1 is *real*, so that $\eta_\rho = 0$. Thus the critical point has an emergent mirror symmetry even if the bare parameters in the model do not. Moreover, the flow out of the fixed point along the single unstable direction is also along a line with v_1 real. Thus the crossover between the intermediate fixed point and the trivial fixed point, which determines the crossover scaling function, also has emergent mirror symmetry. Mirror symmetry breaking is an irrelevant perturbation at the critical fixed point.

If $\epsilon_\sigma = 3\epsilon_\rho$ then the lattice vectors have a triangular symmetry. In this case, the fixed point is at $v_1 = v_2 = \epsilon_\rho$. This means that the periodic potential at the fixed point has emergent *triangular symmetry* even when the bare potential does not. The unstable flow out of the fixed point is also along the high symmetry line $v_1 = v_2$.

It seems quite likely that the critical fixed point and unstable flows connecting it to the trivial fixed points retain their high symmetry even outside the perturbative small ϵ regime. This suggests that in general the critical fixed point has mirror symmetry and that at $g = 1/\sqrt{3}$ it has triangular symmetry. We will use this fact below to determine the critical conductance at $g = 1/\sqrt{3}$.

2. Kubo conductance, mobility, and duality relations

The spin and charge conductances in the Luttinger liquid model computed by the Kubo formula are given by a retarded current-current correlation function. For the present discussion it is useful to write this as an imaginary time correlation function, which can be analytically continued to real time via $i\omega \to \omega + i\eta$ before taking the $\omega \to 0$ limit. Then

$$G_{\alpha\beta}^K(i\omega_n) = \frac{1}{\hbar|\omega_n|}\int d\tau e^{i\omega_n\tau}\langle J_\alpha(\tau)J_\beta(0)\rangle, \tag{3.19}$$

where the spin and charge currents are $J_\alpha = e\partial_t\theta_\alpha/\pi = e[\theta, \mathcal{H}]/(i\pi\hbar)$. This may be expressed as

$$G_{\alpha\beta}^K(\omega_n) = 2\frac{e^2}{h}\sqrt{g_\alpha g_\beta}\mu_{\alpha\beta}(\omega_n), \tag{3.20}$$

where the *mobility* of the QBM model is

$$\mu_{\alpha\beta}(\omega_n) = 2\pi|\omega_n|\langle r_\alpha(-\omega_n)r_\beta(\omega_n)\rangle. \tag{3.21}$$

$\mu_{\alpha\beta}$ is normalized so that when $v_\mathbf{G} = 0$ $\mu_{\alpha\beta} = \delta_{\alpha\beta}$.

The conductance or equivalently $\mu_{\alpha\beta}$ can also be computed from the dual model. It is given by
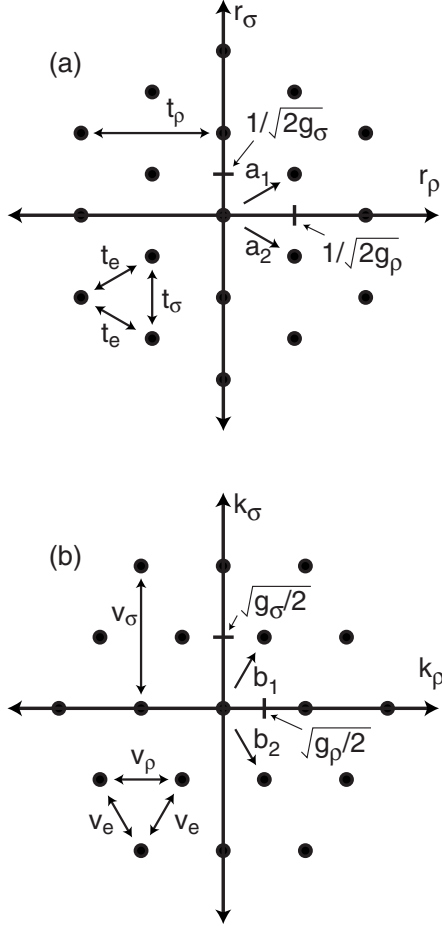
FIG. 8. (a) Minima of the periodic potential $V(\mathbf{r})$ in Eq. (3.13). (b) Minima of $V(\mathbf{k})$ in dual theory (3.16). When $g_\sigma = 3g_\rho$ both periodic potentials have triangular symmetry at the critical point, which implies that the mobility $\mu_{\alpha\beta}^*$ is isotropic. This occurs at $g = 1/\sqrt{3}$.

$$\mu_{\alpha\beta} = \delta_{\alpha\beta} - \tilde{\mu}_{\alpha\beta}, \tag{3.22}$$

where the *dual* mobility is

$$\tilde{\mu}_{\alpha\beta}(\omega_n) = 2\pi|\omega_n|\langle k_\alpha(-\omega_n)k_\beta(\omega_n)\rangle. \tag{3.23}$$

Equations (3.22) and (3.23) are obvious in the perfectly transmitting and perfectly reflecting limits. They can be derived more generally by starting with a Hamiltonian formulation of the action, analogous to Eq. (2.13), which involves both $\mathbf{r}$ and $\mathbf{k}$. $\mu_{\alpha\beta}$ can then be computed either by first integrating out $\mathbf{k}$ to obtain Eq. (3.21) or first integrating out $\mathbf{r}$ to obtain Eq. (3.23).

Since $g_\rho = 1/g_\sigma = g$, the dual theory depicted in Fig. 8(b) is identical to the original theory shown in Fig. 8(a) with the identification $r_\rho \leftrightarrow k_\sigma$, $r_\sigma \leftrightarrow k_\rho$. It follows that the mobility $\mu_{\alpha\beta}^*$ of the fixed point satisfies

$$\mu_{\alpha\beta}^* = [\sigma^x \tilde{\mu}^* \sigma^x]_{\alpha\beta}. \tag{3.24}$$

In addition, if u parametrizes the relevant direction at the critical fixed point, then under the duality $u \to -u$. It follows that slightly away from the critical fixed point we have

$$\mu_{\alpha\beta}(u) = [\sigma^x \tilde{\mu}(-u)\sigma^x]_{\alpha\beta}. \tag{3.25}$$

Properties (3.22) and (3.25) imply that $\mu_{\rho\rho}(u) = 1 - \mu_{\sigma\sigma}(-u)$. Using Eqs. (3.2), (3.20), and (A15), this leads directly to property (3.3) of the crossover scaling function.

An additional set of relations follows from the equivalence between the theory characterized by g and the dual theory characterized by $1/g$. From this we conclude that

$$\mu_{g,\alpha\beta}(u) = \tilde{\mu}_{1/g,\alpha\beta}(u). \tag{3.26}$$

This, combined with Eqs. (3.2), (3.20), (3.22), and (A15) leads to Eq. (3.4).

3. Conductance at $g = 1/\sqrt{3}$

When $g = 1/\sqrt{3}$ the lattice generated by $\mathbf{b}_1$ and $\mathbf{b}_2$ has triangular symmetry. In Sec. III B 1 we argued that this means that at the critical fixed point the periodic potential also has triangular symmetry. The C_6 rotational symmetry of the triangular lattice requires that the mobility is *isotropic*,

$$\mu_{\alpha\beta} = \mu_0 \delta_{\alpha\beta}. \tag{3.27}$$

Combining Eqs. (3.22), (3.24), and (3.27) requires that

$$\mu_0 = \frac{1}{2}. \tag{3.28}$$

It follows from Eq. (3.20) that the Kubo formula spin and charge conductances are given by

$$G_{\rho\rho}^K = \sqrt{3}\frac{e^2}{h}, \quad G_{\sigma\sigma}^K = \frac{1}{\sqrt{3}}\frac{e^2}{h}. \tag{3.29}$$

It is well known that the physical conductance measured with leads is not given by the Kubo conductance.[35–39] Rather, the Kubo conductance needs to be modified to account for the contact resistance between the Luttinger liquid and the leads. In the Appendix we review the relation between the physical four-terminal conductance and the Kubo conductance. From Eq. (A19) we conclude that

$$G_{XX} = G_{YY} = (\sqrt{3} - 1)\frac{e^2}{h}. \tag{3.30}$$

C. Weak interactions: $g = 1 - \epsilon$

In this section we develop a perturbative expansion for weak interactions to compute exactly the crossover scaling function $\mathcal{G}_g(X)$ as well as the critical exponent α_g for $g = 1 - \epsilon$. A similar approach was employed by Matveev *et al.*[27] to compute the scaling function for the crossover between the weak barrier and strong barrier limits in a single-channel Luttinger liquid. In the single-channel problem the transmission for noninteracting electrons is characterized by a transmission probability $\mathcal{T}$. Weak forward scattering interactions lead to an exchange correction to $\mathcal{T}$ at first order in the interactions. This correction diverges for $E \to E_F$ as $\ln|E - E_F|$. Matveev *et al.*[27] used a renormalization group argument to sum the logarithmic divergent corrections to all orders to obtain the exact transmission $\mathcal{T}(E)$.

For noninteracting electrons, the QSHI point contact is characterized by a 4×4 scattering matrix S_{ij} which relates the incoming wave in lead i to the outgoing wave in lead j,

$$|\psi_{i,\text{out}}\rangle = S_{ij}|\psi_{j,\text{in}}\rangle. \tag{3.31}$$

In terms of S_{ij} the four-terminal conductance is

$$G_{ij} = \frac{e^2}{h}(\delta_{ij} - |S_{ij}|^2). \tag{3.32}$$

Under time reversal $\Theta|\psi_{i,\text{out(in)}}\rangle = +(-)Q_{ij}|\psi_{j,\text{in(out)}}\rangle$, where $Q = \text{diag}(1,-1,1,-1)$. This leads to the constraint $S = -QS^TQ$. This combined with unitarity $S^\dagger S = 1$ allows S to be parametrized as

$$S = U^\dagger \begin{pmatrix} 0 & t & f & r \\ t & 0 & r^* & -f^* \\ -f & r^* & 0 & -t^* \\ r & f^* & -t^* & 0 \end{pmatrix} U, \tag{3.33}$$

where $U_{ij} = \delta_{ij}e^{i\chi_i}$ is an unimportant gauge transformation. The complex numbers t and r describe the amplitudes for spin conserving transmission and reflection across the point contact, while f describes the amplitude for tunneling across the junction, combined with a spin flip. $f = 0$ if spin is conserved. The conductance can be expressed in terms of the transmission probabilities $\mathcal{R} = |r|^2$, $\mathcal{T} = |t|^2$, and $\mathcal{F} = |f|^2$, which satisfy $\mathcal{R} + \mathcal{T} + \mathcal{F} = 1$. We find

$$G_{XX} = \frac{2e^2}{h}(\mathcal{T} + \mathcal{F}),$$

$$G_{YY} = \frac{2e^2}{h}(\mathcal{R} + \mathcal{F}),$$

$$G_{ZZ} = \frac{2e^2}{h}(1 - \mathcal{F}),$$

$$G_{AB} = 0 \quad \text{for } A \neq B. \tag{3.34}$$

For a generic four-terminal conductance device time-reversal symmetry guarantees only the reciprocity relation[40] $G_{ij} = G_{ji}$ (or equivalently $G_{AB} = G_{BA}$). For the QSHI point contact, the spin filtered nature of the edge states leads to additional constraints. First, the amplitude for an electron to be reflected back into the lead it came from is $S_{ii} = 0$. Thus $G_{ii} = e^2/h$. A second less obvious constraint is that $G_{13} = G_{24}$, which when combined with reciprocity and unitarity is equivalent to $G_{12} = G_{34}$ and $G_{14} = G_{23}$. This leads to the vanishing of the skew conductance G_{XY}, as well as G_{XZ} and G_{YZ} even when mirror symmetries $\mathcal{M}_X$ and $\mathcal{M}_Y$ are explicitly violated. This is a property of the noninteracting electron model and can be violated with electron electron interactions if the mirror symmetries are absent.

In order to compute the renormalization of the S matrix due to interactions it is useful to study the perturbative expansion of the single electron thermal Green's function, which can be represented as a matrix in the lead indices i, j as well as the channel labels $a = \text{in/out}$. Here,
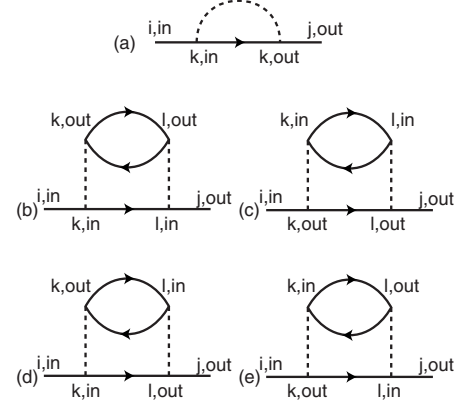
FIG. 9. Feynman diagrams for the single electron Green's function. The dashed line is the interaction $u_2(\psi_{\text{in}}^\dagger\psi_{\text{in}})(\psi_{\text{out}}^\dagger\psi_{\text{out}})$. The exchange diagram (a) vanishes because it involves S_{kk} and diagrams (b) and (c) cancel one another. (d) and (e) lead to a logarithmic correction to the S matrix given in Eq. (3.38).

$$\mathbb{G}_{ij}^{ab}(x,\tau;x',\tau') = -i\langle T_\tau[\psi_{i,a}(x,\tau)\psi_{j,b}^\dagger(x',\tau')]\rangle, \tag{3.35}$$

where T_τ denotes imaginary time ordering. For noninteracting electrons we have

$$\mathbb{G}_{ij}(z,z') = \frac{1}{2\pi i}\begin{pmatrix} \dfrac{\delta_{ij}}{z-z'} & \dfrac{S_{ji}^*}{z-\bar{z}'} \\ \dfrac{S_{ij}}{\bar{z}-z'} & \dfrac{\delta_{ij}}{\bar{z}-\bar{z}'} \end{pmatrix}, \tag{3.36}$$

where $z = \tau + ix$ and $\bar{z} = \tau - ix$, and the $a = \text{in/out}$ indices are displayed in matrix form.

We now compute the perturbative corrections to $\mathbb{G}_{ij}^{\text{out,in}}$ using the standard diagrammatic technique. For simplicity, we adopt a model in which $u_4 = 0$, so that the only interaction term involves $u_2(\psi_{\text{in}}^\dagger\psi_{\text{in}})(\psi_{\text{out}}^\dagger\psi_{\text{out}})$. This considerably simplifies the analysis because many of the diagrams are zero. For instance, the exchange diagram shown in Fig. 9(a), which was responsible for the renormalization in the single-channel Luttinger liquid problem, is zero because it must involve $\mathbb{G}_{kk}^{\text{in,out}}$. This off-diagonal Green's function depends on S_{kk} which is zero due to the time-reversal-symmetry constraint. From Eq. (2.9), $g = \sqrt{(2\pi v_F - \lambda_2)/(2\pi v_F + \lambda_2)} \sim 1 - \lambda_2/(2\pi v_F)$. Thus for $g = 1 - \epsilon$ we may replace u_2 by $2\pi v_F \epsilon$. The nonzero diagrams at second order in u_2 are shown in Figs. 9(b)–9(e). Evaluating the second-order diagrams gives a Green's function of the form

$$\mathbb{G}_{ij}^{\text{out,in}} = \frac{1}{2\pi i}\frac{S_{ij}'}{\bar{z}-z'} \tag{3.37}$$

with

$$S_{ij}' = S_{ij} + \frac{\epsilon^2}{4}\ln\frac{\Lambda}{E}\left[S_{ij}S_{ji}S_{ji}^* - \sum_{kl} S_{ik}S_{kl}S_{lk}^*S_{kl}^*S_{lj}\right], \tag{3.38}$$

where Λ are E are ultraviolet and infrared cutoffs, respectively. The first term in the brackets was due to the diagram
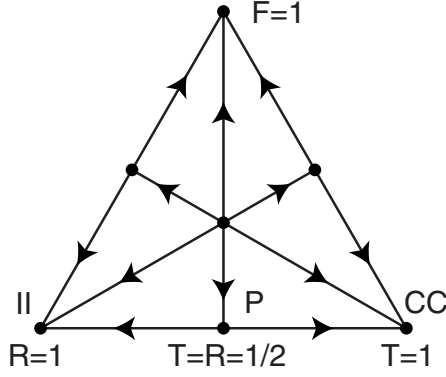
FIG. 10. Renormalization group flow diagram for the transmission probabilities $\mathcal{T}$, $\mathcal{R}$, and $\mathcal{F}$ based on Eq. (3.40) represented in a ternary plot. The CC, II, and P fixed points of interest in this paper, which have $\mathcal{F}=0$ are on the bottom of the triangle.

in Fig. 9(d), while the second term was from Fig. 9(e). Diagrams in Figs. 9(b) and 9(c) canceled each other. Rescaling the cutoff $\Lambda \to \Lambda e^{-\ell}$ leads to a renormalization group flow equation for S_{ij},

$$\frac{dS_{ij}}{d\ell} = \frac{\epsilon^2}{4}\left[S_{ij}S_{ji}S_{ji}^* - \sum_{kl} S_{ik}S_{kl}S_{lk}^*S_{kl}^*S_{lj} \right]. \qquad (3.39)$$

It is useful to rewrite this in terms of the transmission probabilities $\mathcal{T}$, $\mathcal{R}$, $\mathcal{F}$. The renormalization group flow equation then can be written in the form

$$d\mathcal{T}/d\ell = \epsilon^2 \mathcal{T}(\mathcal{T} - \mathcal{T}^2 - \mathcal{R}^2 - \mathcal{F}^2),$$

$$d\mathcal{R}/d\ell = \epsilon^2 \mathcal{R}(\mathcal{R} - \mathcal{T}^2 - \mathcal{R}^2 - \mathcal{F}^2),$$

$$d\mathcal{F}/d\ell = \epsilon^2 \mathcal{F}(\mathcal{F} - \mathcal{T}^2 - \mathcal{R}^2 - \mathcal{F}^2). \qquad (3.40)$$

The flow diagram as a function of $\mathcal{R}$, $\mathcal{T}$, and $\mathcal{F}$ is shown in Fig. 10. There are seven fixed points. The bottom corners of the triangle are the stable fixed points at $\mathcal{R}=1$, $\mathcal{T}=\mathcal{F}=0$ (the II phase) and $\mathcal{T}=1$, $\mathcal{R}=\mathcal{F}=0$ (the CC phase). The third stable fixed point at the top of the triangle with $\mathcal{F}=1$, $\mathcal{T}=\mathcal{R}=0$ corresponds to the case where an incident electron is transmitted perfectly with a spin flip. This is presumably difficult to access physically. On the edges of the triangle are unstable fixed points describing transitions between the different stable phases. The critical fixed point P of interest in this paper is the one on the bottom of the triangle at $\mathcal{R}=\mathcal{T}=1/2$, $\mathcal{F}=0$. Note that at this fixed point the spin nonconserving spin-orbit processes, represented by $\mathcal{F}$, are irrelevant. At the center of the triangle, at $\mathcal{R}=\mathcal{T}=\mathcal{F}=1/3$ is an unstable fixed point describing a multicritical point.

To describe the critical fixed point P and the crossover to the II and CC phases we now specialize to $\mathcal{F}=0$ and consider the flow equation for the single parameter $\mathcal{T}$ characterizing the point contact

$$d\mathcal{T}/d\ell = -\epsilon^2 \mathcal{T}(1-\mathcal{T})(1-2\mathcal{T}). \qquad (3.41)$$

Equation (3.41) can be integrated to determine the crossover scaling function. If at $\ell=0$, $\mathcal{T}=\mathcal{T}^0$, then,

$$\mathcal{T}(\ell) = \frac{1}{2}\left[1 + \frac{\mathcal{T}^0 - 1/2}{\sqrt{(\mathcal{T}^0 - 1/2)^2 + \mathcal{T}^0(1 - \mathcal{T}^0)e^{-\epsilon^2\ell}}} \right]. \qquad (3.42)$$

As the gate voltage V_G is adjusted through the pinch-off transition, $\mathcal{T}^0$ passes through $1/2$ at $V_G = V_G^*$, so $\mathcal{T}^0 - 1/2 \propto \Delta V_G$. At temperature T we cut off the renormalization group flow at $\Lambda e^{-\ell} \propto T$. The conductance is then given by $G_{XX} = 2(e^2/h)\mathcal{T}[\ell = \ln(\Lambda/T)]$. For $\Delta V_G, T \to 0$ we define $X = (2\mathcal{T}^0 - 1)e^{\epsilon^2\ell/2} \propto \Delta V_G/T^{\epsilon^2/2}$ and write the conductance in the scaling form,

$$G_{XX}(\Delta V_G, T) = 2\frac{e^2}{h}\mathcal{G}_{1-\epsilon}\left(c\frac{\Delta V_G}{T^{\alpha_g}} \right), \qquad (3.43)$$

where c is a nonuniversal constant, the critical exponent is

$$\alpha_{1-\epsilon} = \epsilon^2/2, \qquad (3.44)$$

and

$$\mathcal{G}_{1-\epsilon}(X) = \frac{1}{2}\left[1 + \frac{X}{\sqrt{1+X^2}} \right]. \qquad (3.45)$$

We find that the logarithmic renormalization to the S matrix accounts for the *only* correction to the conductance to linear order in ϵ. In principle one must consider a "random-phase-approximation-type" diagram for the conductance evaluated by the Kubo formula. While this gives a correction for an *infinite* Luttinger liquid at finite frequency, the correction is zero for a finite Luttinger liquid connected to leads in the $\omega \to 0$ limit.[35–38] Since the critical conductance satisfies $G_g^* = 1 - G_{1/g}^*$ it follows that $G_{1-\epsilon}^* = 1/2 + O(\epsilon^3)$.

D. $g = 1/2 + \epsilon$

$g = 1/2$ is at the boundary where the CC and II phases become unstable and the IC phase becomes stable. We will show that when $g = 1/2 + \epsilon$ the critical fixed point describing the transition between the CC and II phases approaches the IC fixed point and can be accessed perturbatively using theory developed in Sec. II B 3. In addition, when $g = 1/2$, the marginal operators $v_\rho \cos 2\theta_\rho$ at the CC fixed point and $\tilde{t} \cos \tilde{\theta}_\rho$ at the IC fixed point can be expressed in terms of fictitious fermion operators. This fermionization process allows the entire crossover between the CC and IC phases to be described using a noninteracting fermion Hamiltonian. A similar fermionization procedure can be used to describe the crossover between the II and IC phases, which connect the marginal operators $\tilde{v}_\sigma \cos \theta_\sigma$ and $\tilde{t}_\sigma \cos 2\tilde{\theta}_\sigma$. This will allow us to compute the full crossover scaling function $\mathcal{G}_g(X)$ for $g = 1/2 + \epsilon$.

We will begin by discussing the perturbative analysis of the IC fixed point and then go on to describe the fermionization procedure.

1. Perturbative analysis

The IC fixed point is described by Eqs. (2.32) and (2.33). When $g = 1/2 + \epsilon$ the perturbations $\tilde{t}_\rho \tau^x \cos \tilde{\theta}_\rho$ and $\tilde{t}_\sigma \tau^z \cos \theta_\sigma$

both have scaling dimension $\Delta = 1 - 2\epsilon$, so the IC fixed point is weakly unstable. When $\widetilde{v}_\sigma = 0$, nonzero $\widetilde{t}_\rho$ is expected to drive the system to the CC phase, while for $\widetilde{t}_\rho = 0$ nonzero $\widetilde{v}_\sigma$ will drive the system to the II phase. Thus, when both $\widetilde{t}_\rho$ and $\widetilde{v}_\sigma$ are nonzero there must be an unstable fixed point which separates the two alternatives. This fixed point can be described by considering the renormalization group flow equations to *third* order in $\widetilde{v}_\sigma$ and $\widetilde{t}_\rho$.

The first-order renormalization group equation for $\widetilde{t}_\rho$ is determined by the scaling dimension $\Delta(\widetilde{t}_\rho)$. The next nonzero term occurs at order $t_\rho v_\sigma^2$. To compute this term it is sufficient to use the theory at $\epsilon = 0$. Consider the third-order term in the cumulant expansion of the partition function when fast degrees of freedom integrated out,

$$\frac{1}{2} \int d\tau_1 d\tau_2 \{ \langle T_\tau [O_\rho(\tau) O_\sigma(\tau_1) O_\sigma(\tau_2)] \rangle_> $$
$$- \langle O_\rho(\tau) \rangle_> \langle T_\tau [O_\sigma(\tau_1) O_\sigma(\tau_2)] \rangle_> \}. \quad (3.46)$$

Here $O_\rho = (\widetilde{t}_\rho / \tau_c) \tau^x \cos \widetilde{\theta}_\rho$ and $O_\sigma = (\widetilde{v}_\sigma / \tau_c) \tau^z \cos \theta_\sigma$. T_τ indicates time ordering, and $\langle \cdot \rangle_>$ denotes a trace over degrees of freedom with $\Lambda / b < \omega < \Lambda$, and we assume for simplicity $b \gg 1$. Since $\widetilde{\theta}_\rho$ and θ_σ are independent and commute with one another the other disconnected terms all cancel. Moreover, the two terms in Eq. (3.46) will cancel each other unless the time ordering of the τ^x and τ^z operators leads to a relative minus sign between them,

$$\langle T_\tau [O_\rho(\tau) O_\sigma(\tau_1) O_\sigma(\tau_2)] \rangle_> $$
$$= s_\pm \langle O_\rho(\tau) \rangle_> \langle T_\tau [O_\sigma(\tau_1) O_\sigma(\tau_2)] \rangle_>, \quad (3.47)$$

where $s_\pm = \mathrm{sgn}(\tau - \tau_1)(\tau - \tau_2)$. Thus the pseudospin operators in Eq. (2.33) play a crucial role in the renormalization of $\widetilde{t}_\rho$. Using the fact that $\langle T_\tau [O_\sigma(\tau_1) O_\sigma(\tau_2)] \rangle_> = \widetilde{v}_\sigma^2 / 2(\tau_1 - \tau_2)^2$ for $\epsilon = 0$ we find that the third-order correction to $\widetilde{t}_\rho$ is $\delta \widetilde{t}_\rho = -t_\rho v_\sigma^2 \ln b$. This leads to the renormalization group flow equation for $\widetilde{t}_\rho$ along with a corresponding equation for $\widetilde{v}_\sigma$,

$$d\widetilde{t}_\rho / d\ell = 2\epsilon \widetilde{t}_\rho - \widetilde{t}_\rho \widetilde{v}_\sigma^2,$$

$$d\widetilde{v}_\sigma / d\ell = 2\epsilon \widetilde{v}_\sigma - \widetilde{v}_\sigma \widetilde{t}_\rho^2. \quad (3.48)$$

The renormalization group flow diagram is shown in Fig. 11. There is an unstable fixed point P at $\widetilde{t}_\rho = \widetilde{v}_\sigma = \sqrt{2\epsilon}$, with a single relevant operator. P separates the flows to the CC and II phases for which $\widetilde{t}_\rho$ or $\widetilde{v}_\sigma$ grows. Note that spin-orbit terms such as v_{so} and $\widetilde{v}_{\mathrm{sf}}$ discussed in Sec. II B 5 are irrelevant at P [see Eq. (2.43)]. This perturbative calculation provides further evidence that P exhibits emergent spin conservation, as well as emergent mirror symmetry. The critical exponent associate with the single relevant operator P is

$$\alpha_{1/2+\epsilon} = 4\epsilon. \quad (3.49)$$

The Kubo conductance $G_{\rho\rho}^K$ at the fixed point can be computed from Eq. (3.19) by identifying the current operator

$$I_\rho = (\widetilde{t}_\rho / \tau_c) \sin \widetilde{\theta}_\rho. \quad (3.50)$$

This leads to
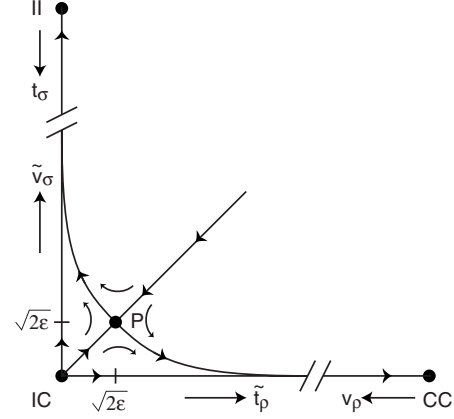
FIG. 11. Renormalization group flow diagram characterizing the critical fixed point P for $g = 1/2 + \epsilon$. When $\widetilde{v}_\sigma$ and $\widetilde{t}_\rho$ are small, the flows are given by Eq. (3.48). On the axis $\widetilde{v}_\sigma = 0$ the fermionization procedure outlined in Sec. III D 2 determines the entire crossover between the IC and CC fixed points. A similar theory describes the crossover between the IC and II fixed points for $\widetilde{t}_\rho = 0$.

$$G_{\rho\rho}^K = \frac{e^2}{h} \pi^2 \widetilde{t}_\rho^2. \quad (3.51)$$

It is useful to define $\mathcal{T}_\rho = \pi^2 \widetilde{t}_\rho^2$. We will see in Sec. III D 2 that this can be interpreted as a transmission probability for fictitious free fermions that describe the problem at $g = 1/2$. In terms of $\mathcal{T}_\rho$ (noting that $\mathcal{T}_\rho \ll 1$ in this perturbative regime) we may use Eq. (A19) to write the physical conductance as

$$G_{XX} = \frac{e^2}{h} \mathcal{T}_\rho. \quad (3.52)$$

A similar calculation gives

$$G_{YY} = \frac{e^2}{h} \mathcal{R}_\sigma, \quad (3.53)$$

where $\mathcal{R}_\sigma = \pi^2 \widetilde{v}_\sigma^2$ can similarly be interpreted as a reflection probability for a different fictitious free fermion at $g = 1/2$. At the critical fixed point $\mathcal{T}_\rho = \mathcal{R}_\sigma = 2\pi^2 \epsilon$. Thus,

$$G_{XX}^* = G_{YY}^* = 2 \frac{e^2}{h} \pi^2 \epsilon,$$

$$G_{XY}^* = 0. \quad (3.54)$$

The behavior away from the critical point can be determined by integrating Eq. (3.48). To this end it is helpful to rewrite Eq. (3.48) in terms of $\mathcal{T}_\rho$ and $\mathcal{R}_\sigma$ in the following forms:

$$d(\mathcal{T}_\rho - \mathcal{R}_\sigma) / d\ell = 4\epsilon (\mathcal{T}_\rho - \mathcal{R}_\sigma),$$

$$d \ln(\mathcal{T}_\rho / \mathcal{R}_\sigma) / d\ell = (2/\pi^2)(\mathcal{T}_\rho - \mathcal{R}_\sigma). \quad (3.55)$$

If $(\mathcal{T}_\rho, \mathcal{R}_\sigma) = (\mathcal{T}_\rho^0, \mathcal{R}_\sigma^0)$ for $\ell = 0$, then we find

$$\mathcal{T}_\rho(\ell) = \frac{(\mathcal{T}_\rho^0 - \mathcal{R}_\sigma^0)e^{4\epsilon\ell}}{1 - \frac{\mathcal{R}_\sigma^0}{\mathcal{T}_\rho^0}\exp\left[-\frac{\mathcal{T}_\rho^0 - \mathcal{R}_\sigma^0}{2\pi^2\epsilon}(e^{4\epsilon\ell} - 1)\right]},$$

$$\mathcal{R}_\sigma(\ell) = \frac{(\mathcal{R}_\sigma^0 - \mathcal{T}_\rho^0)e^{4\epsilon\ell}}{1 - \frac{\mathcal{T}_\rho^0}{\mathcal{R}_\sigma^0}\exp\left[-\frac{\mathcal{R}_\sigma^0 - \mathcal{T}_\rho^0}{2\pi^2\epsilon}(e^{4\epsilon\ell} - 1)\right]}. \quad (3.56)$$

At the pinch-off transition $V_G = V_G^*$, $\mathcal{R}_0 = \mathcal{T}_0$. Thus, $\mathcal{T}_0 - \mathcal{R}_0 \propto \Delta V_G$. At temperature T we cut off the renormalization group flow at $\Lambda e^{-\ell} \propto T$. Thus, in the limit $\Delta V_G, T \to 0$ we define $X = (\mathcal{T}_\rho^0 - \mathcal{R}_\sigma^0)e^{4\epsilon\ell}/2 \propto \Delta V_G/T^{4\epsilon}$. The conductance then has the form

$$G_{XX}(\Delta V_G, T) = 2\frac{e^2}{h}\mathcal{G}_g\left(c\frac{\Delta V_G}{T^{\alpha_g}}\right),$$

$$G_{YY}(\Delta V_G, T) = 2\frac{e^2}{h}\mathcal{G}_g\left(-c\frac{\Delta V_G}{T^{\alpha_g}}\right), \quad (3.57)$$

with

$$\mathcal{G}_{1/2+\epsilon}(X) = \frac{X}{1 - e^{-X/(\pi^2\epsilon)}}. \quad (3.58)$$

This perturbative calculation is only valid when $\mathcal{T}_\rho, \mathcal{R}_\sigma \ll 1$. Thus Eq. (3.58) breaks down at low temperature since as the energy is lowered either $\mathcal{T}_\rho$ or $\mathcal{R}_\sigma$ grows. Equation (3.58) is valid as long as $|X| \ll 1$. Note, however, that when $\epsilon \ll 1$ we have $\mathcal{G}_{1/2+\epsilon}(X) = X\theta(X)$ when $\epsilon \ll X \ll 1$. In this regime, the smaller of $\mathcal{T}$ and $\mathcal{R}$ has gone to zero. Thus we have

$$\left.\begin{array}{l}\mathcal{T}_\rho(\ell) = (\mathcal{T}_\rho^0 - \mathcal{R}_\sigma^0)e^{4\epsilon\ell} \\ \mathcal{R}_\sigma(\ell) = 0\end{array}\right\}(\mathcal{T}_\rho^0 - \mathcal{R}_\sigma^0) > 0,$$

$$\left.\begin{array}{l}\mathcal{T}_\rho(\ell) = 0 \\ \mathcal{R}_\sigma(\ell) = (\mathcal{R}_\sigma^0 - \mathcal{T}_\rho^0)e^{4\epsilon\ell}\end{array}\right\}(\mathcal{R}_\sigma^0 - \mathcal{T}_\rho^0) > 0, \quad (3.59)$$

and the unstable flow is either on the x or y axis of Fig. 11. In Sec. III D 2 we will solve the crossover exactly on these lines. This will allow us to compute the $\mathcal{G}_{1/2+\epsilon}(X)$ exactly for all X.

2. Fermionization

In this section we study the crossover between the IC fixed point and the CC and II fixed points for $g = 1/2 + \epsilon$. There are two cases to consider. First, for $\Delta V_G > 0$ we will study the crossover between the IC and CC on the horizontal axis of Fig. 11 with $\tilde{v}_\sigma = 0$. This problem can be mapped to a single-channel one-dimensional Fermi gas with weak electron-electron interactions proportional to ϵ. This allows us to use the method of Matveev *et al.*[27] to compute the crossover scaling functions $\mathcal{G}_{g,XX}(X)$ and $\mathcal{G}_{g,YY}(X)$ for $X > 0$ exactly. For $\Delta V_G < 0$ the crossover between the IC and II fixed points is on the vertical axis of Fig. 11 with $\tilde{t}_\rho = 0$. This can be fermionized by introducing a *different* set of free fermions to compute the scaling functions for $X < 0$. The latter

calculation (which is virtually identical to the former) is unnecessary, however, because we can use Eq. (3.3) to deduce the scaling functions for $X < 0$. We will therefore focus on the IC to CC crossover.

The crossover between the IC and the CC fixed points can be described by the action in the CC limit,

$$S_{CC} = \frac{1}{\beta}\sum_n \frac{1}{2\pi g}|\omega_n||\theta_\rho(\omega_n)|^2 + \int d\tau v_\rho \cos 2\theta_\rho. \quad (3.60)$$

$v_\rho \ll 1$ describes the CC phase. When $v_\rho \gg 1$ the dual theory, formulated as in Sec. II B 3 in terms of instantons with amplitude $\tilde{t}_\rho$, describes the IC phase. When $\tilde{v}_\sigma = 0$ at the IC fixed point we can safely ignore the pseudospin and set $\tau^x = 1$.

For $g = 1/2$ this model is equivalent to the bosonized representation of a weak link in a single-channel *noninteracting* fermion with weak backscattering. Here,

$$\mathcal{H}_f = -iv\tilde{\psi}^\dagger\partial_x\sigma^z\tilde{\psi} + v_f\tilde{\psi}^\dagger\sigma^x\tilde{\psi}\delta(x), \quad (3.61)$$

where $\tilde{\psi} = (\tilde{\psi}_R, \tilde{\psi}_L)^T$ is a two component fermion operator describing right and left movers. Using bosonization relation (2.3) we identify $2\theta_\rho = \phi_R - \phi_L$ and $v_f = \pi v_\rho/v$. The free fermion problem is solvable and characterized by a transmission probability $\mathcal{T}_\rho = \mathrm{sech}^2(v_f/v)$. The free fermion solution therefore connects the CC limit ($\mathcal{T}_\rho = 1$) with the IC limit ($\mathcal{T}_\rho = 0$).

The Kubo conductance $G_{\rho\rho}^K$ may be computed with the identification $J_\rho = \partial_t\theta_\rho/\pi = v\tilde{\psi}^\dagger\sigma^z\tilde{\psi}$, giving

$$G_{\rho\rho}^K = \frac{e^2}{h}\mathcal{T}_\rho. \quad (3.62)$$

Note that this is the same as Eq. (3.52), derived in the opposite limit near the IC fixed point. When v_ρ is large, $\mathcal{T}_\rho \ll 1$, and we can identify $\mathcal{T}_\rho = (\pi\tilde{t}_\rho)^2$. The physical conductance, measured with leads, can be determined following the analysis in the Appendix. From Eq. (A19) we find

$$G_{XX} = 2\frac{e^2}{h}\frac{\mathcal{T}_\rho}{2 - \mathcal{T}_\rho}. \quad (3.63)$$

Since $v_\sigma = 0$ in Eq. (3.60), we have

$$G_{YY} = 0. \quad (3.64)$$

For $g = 1/2 + \epsilon$ the IC fixed point becomes slightly unstable, while the CC fixed point becomes slightly stable. In this case the free fermion problem includes a weak *attractive* interaction

$$\mathcal{H}_f^{int} = -u_f(\tilde{\psi}_L^\dagger\tilde{\psi}_L)(\tilde{\psi}_R^\dagger\tilde{\psi}_R), \quad (3.65)$$

with $u_f = 2\pi v\epsilon$. This leads to a logarithmic renormalization of $\mathcal{T}_\rho$, which drives a crossover to the CC limit. The correction to $\mathcal{T}_\rho$ occurs at *first* order in u_f and is due to the exchange diagram, shown in Fig. 9(a). The analysis is exactly the same as that performed by Matveev *et al.* As in Sec. III C the result can be cast in terms of a renormalization group flow equation for $\mathcal{T}_\rho$,

$$dT_\rho/d\ell = 4\epsilon T_\rho(1 - T_\rho). \tag{3.66}$$

Integrating Eq. (3.66) gives

$$T_\rho(\ell) = \frac{T_\rho^0 e^{4\epsilon\ell}}{1 + T_\rho^0(e^{4\epsilon\ell} - 1)}, \tag{3.67}$$

where $T_\rho^0 = T_\rho(\ell=0)$. The scaling function for $\Delta V_G > 0$ then follows by using the initial condition from Eq. (3.59), so that $T_\rho^0 \propto \Delta V_G$. Then, for $\Delta V_G, T \to 0$ we define $X = T_\rho^0 e^{4\epsilon\ell}/2 \propto \Delta V_G/T^{4\epsilon}$. Using Eqs. (3.63), (3.64), and (3.67), the conductance has the scaling form for $X > 0$,

$$\mathcal{G}_{XX,1/2+\epsilon}(X) = \frac{X}{X + 1},$$

$$\mathcal{G}_{YY,1/2+\epsilon}(X) = 0. \tag{3.68}$$

Using Eq. (3.3), we may deduce the corresponding behavior for $\Delta V_G < 0$ (or $X < 0$). The scaling function then has the form

$$\mathcal{G}_{1/2+\epsilon}(X) = \theta(X)\frac{X}{X + 1}. \tag{3.69}$$

Note that for $X \ll 1$, $\mathcal{G}_{1/2+\epsilon}(X) = X\theta(X)$, in agreement with the limiting behavior of Eq. (3.58) for $X \gg \epsilon$. These two expressions can thus be combined to give

$$\mathcal{G}_{1/2+\epsilon}(X) = \frac{X}{X + 1 - e^{-X/(\pi^2\epsilon)}}, \tag{3.70}$$

which reproduces Eq. (3.58) when $|X| \sim \epsilon \ll 1$ and Eq. (3.69) when $|X| \gg \epsilon$. This function is plotted in Fig. 7(b). Note, however, that this formula does not correctly capture the leading behavior for $X < 0$ when $|X| \gg \epsilon$. In particular, it misses the $X \to -\infty$ behavior, which Eqs. (3.5) and (3.6) predict is proportional to $|X|^{-1/(8\epsilon)}$. This regime is analyzed in Sec. III D 3.

3. Rebosonization

We now analyze the leading behavior of $\mathcal{G}_{1/2+\epsilon}(X)$ for $X < 0$ and $|X| \gg \epsilon$ when ϵ is small. Equivalently, we consider $\mathcal{G}_{1/2+\epsilon,YY}(X)$ for $X > 0$. This requires extending the renormalization group flow equation for $\tilde{v}_\sigma$ given in Eq. (3.48) to all $\tilde{t}_\rho$ (or equivalently T_ρ). This can be done by using the fermionized representation of $\tilde{t}_\rho\tau^x\cos\tilde{\theta}_\rho$ in Eq. (2.33). The key point is that the presence of the pseudospin operator τ^x means that the operator $\tilde{v}_\sigma\tau^z\cos\theta_\sigma$ changes the sign of the transmission amplitude for the fermions $\tilde{\psi}$. This results in an x-ray edge-like contribution to the renormalization of $\tilde{v}_\sigma$. This can be computed by a method analogous to that used by Schotte and Schotte[41] to solve the x-ray edge problem, which involves transforming the noninteracting fermions to even and odd parity scattering states and then rebosonizing. This approach was used to study the x-ray edge problem in a Luttinger liquid in Ref. 42.

We begin by writing Eq. (2.33), $\mathcal{H} = \mathcal{H}_\sigma + \mathcal{H}_\rho$, with

$$\mathcal{H}_\sigma = \mathcal{H}_\sigma^0 + \tilde{v}_\sigma\tau^z\cos\theta_\sigma \tag{3.71}$$

and

$$\mathcal{H}_\rho = -iv\tilde{\psi}^\dagger\sigma^z\partial_x\tilde{\psi} + t_f\tau^x\tilde{\psi}^\dagger\sigma^x\tilde{\psi}\delta(x). \tag{3.72}$$

Here $\mathcal{H}_\sigma^0$ is the σ part of Eq. (2.13), and we explicitly account for the pseudospin τ^x. Equation (3.72) can be rebosonized by first replacing $\psi_2(x) \to \psi_2(-x)$, which transforms the nonchiral fermions to chiral fermions, eliminating the σ^z in the first term but leaving the second term alone. Then we perform a SU(2) rotation $(\tilde{\psi}_1, \tilde{\psi}_2) \to (\tilde{\psi}_e, \tilde{\psi}_o)$, which changes σ^x in the second term into σ^z. $\tilde{\psi}_{e(o)}$ describe the even (odd) parity scattering states characterized by scattering phase shifts $\delta_e = -\delta_o$ that specify $\tilde{\psi}_{e(o)}(x>0) = e^{2i\delta_{e(o)}}\tilde{\psi}_{e(o)}(x<0)$. We next bosonize $\tilde{\psi}_{e,o} \to e^{i\phi_{e,o}}/\sqrt{2\pi x_c}$ and define $\phi_\pm = \phi_e \pm \phi_o$. Then

$$\mathcal{H}_\rho = \frac{v}{8\pi}[(\partial_x\phi_+)^2 + (\partial_x\phi_-)^2] + \frac{v}{2\pi}\delta_-\tau^x(\partial_x\phi_-)\delta(x), \tag{3.73}$$

where $\phi_\pm$ obey, $[\phi_\pm(x), \phi_\pm(x')] = 2\pi i\,\text{sgn}(x-x')$. $\delta_- = \delta_e - \delta_o$ is related to the transmission probability by

$$T_\rho = \sin^2\delta_-. \tag{3.74}$$

δ_- can be eliminated from Eq. (3.73) by the canonical transformation $U = \exp[i\tau^x\delta_-\phi_-(x=0)/(2\pi)]$, which shifts $\phi_- \to \phi_- + \text{sgn}(x)\delta_-\tau^x$. This transformation also rotates τ^z in Eq. (3.71), which becomes

$$\mathcal{H}_\sigma = H_\sigma^0 + \tilde{v}_\sigma[\tau^+ e^{i\phi_-\delta_-/\pi} + \tau^- e^{-i\phi_-\delta_-/\pi}]\cos\theta_\sigma, \tag{3.75}$$

where $\tau^\pm = \tau^z \pm i\tau^y$. The renormalization of $\tilde{v}_\sigma$ can then easily be determined for arbitrary δ_-. We find

$$\frac{d\tilde{v}_\sigma}{d\ell} = \left(2\epsilon - \left(\frac{\delta_-}{\pi}\right)^2\right)\tilde{v}_\sigma. \tag{3.76}$$

For small $\tilde{t}_\rho$, $\delta_- = \pi\tilde{t}_\rho$, and Eq. (3.76) reproduces Eq. (3.48). However, Eq. (3.76) remains valid to lowest order in ϵ for all T_ρ.

We now integrate Eq. (3.55) to a scale ℓ_0 where from Eq. (3.56) $T_\rho(\ell_0) = 2X_0$ and $\mathcal{R}_\sigma(\ell_0) = 2X_0 e^{-X_0/(\pi^2\epsilon)}$ are small. [Here $X_0 = (T_\rho^0 - \mathcal{R}_\sigma^0)e^{4\epsilon\ell_0}/2$.] We then use that as an initial value for Eq. (3.76), which we integrate assuming $T_\rho(\ell)$ is given by Eq. (3.67) and is unaffected by the small $\mathcal{R}_\sigma$. Expressing Eq. (3.67) in terms of Eq. (3.74) we have

$$\delta_-(\ell) = \tan^{-1}[\delta_-(\ell_0)e^{2\epsilon(\ell-\ell_0)}], \tag{3.77}$$

where $\delta_-(\ell_0) = \sin^{-1}\sqrt{T_\rho(\ell_0)} \sim \sqrt{2X_0}$. As before, we define $X = (T_\rho^0 - \mathcal{R}_\sigma^0)e^{4\epsilon\ell}/2$. We may express $G_{YY} = (e^2/h)\mathcal{R}_\sigma$ with $\mathcal{R}_\sigma = \pi^2\tilde{v}_\sigma^2$. Integrating Eq. (3.76) we then find

$$G_{YY}(X) = 2\frac{e^2}{h}Xe^{-F(X)/\epsilon}, \tag{3.78}$$

where

$$F(X) = \frac{1}{\pi^2}\int_0^{\sqrt{2X}}\frac{dx}{x}(\tan^{-1}x)^2. \tag{3.79}$$

Thus, for $X < 0$, $|X| \gg \epsilon$, and $\epsilon \to 0$ we find

$$\mathcal{G}_{1/2+\epsilon}(X) = |X|e^{-F(|X|)/\epsilon}. \qquad (3.80)$$

The asymptotic behavior $F(X) = X/\pi^2$ for $|X| \ll 1$ reproduces Eq. (3.58) when $|X| \gg \epsilon$. For $|X| \gg 1$ we find

$$F(X \to \infty) = \frac{1}{8}\ln 2X - \frac{7\zeta(3)}{4\pi^2}, \qquad (3.81)$$

where $\zeta(3) = 1.20$ is the Riemann zeta function. This gives the asymptotic behavior

$$\mathcal{G}_{1/2+\epsilon}(X \to -\infty) = \left(\frac{e^{14\zeta(3)/\pi^2}}{2|X|}\right)^{1/8\epsilon}, \qquad (3.82)$$

which is quoted in Table II.

IV. DISCUSSION AND CONCLUSION

In this paper we have examined several properties of a point contact in a QSHI. We showed that the pinch-off as a function of gate voltage is governed by a nontrivial quantum phase transition, which leads to scaling behavior of the conductance as a function of temperature and gate voltage characterized by a universal scaling function. We computed this scaling function and other properties of the critical point in certain solvable limits which provide an overall picture of the behavior as a function of the Luttinger liquid parameter g.

In addition, we showed that the four-terminal conductance has a simple structure when expressed in terms of the natural variables, G_{AB}, and that at the low temperature fixed points, the leading corrections to the different components of G_{AB} can have different temperature dependences. In particular, we showed that the skew conductance G_{XY} vanishes as T^γ with $\gamma \geq 2$.

Finally, we showed that for strong interactions, $g < 1/2$, the stable phase is the time-reversal breaking insulating phase. Transport in that phase occurs via novel fractionalized excitations that have clear signatures in noise correlations.

There are a number of problems for future research that our work raises. We will divide the discussion into experimental and theoretical issues.

A. Experimental issues

The QSHI has been observed in transport experiments on HgTe/HgCdTe quantum well structures. A crucial issue is the value of the interaction parameter g. A simple estimate can be developed based on the long range Coulomb interaction.[43] First consider the limit $\xi \gg w$, where w is the quantum well width and ξ is the evanescent decay length of the edge state wave function into the bulk QSHI. We model the edge state as a two-dimensional charged sheet with a charge density profile proportional to $\theta(x)\exp(-2x/\xi)$, a distance d above a conducting ground plane. The long range interaction then leads to $u_2 = u_4 = (2e^2/\epsilon)\ln(4e^\gamma d/\xi)$, where ϵ is the dielectric constant and $\gamma = 0.577$ is Euler's constant. As a second model, assume $\xi \ll w$ and model the edge state as a uniformly charged two-dimensional strip of width w perpendicular to a

ground plane a distance d away. This gives $u_2 = u_4 = (2e^2/\epsilon)\ln(2e^{3/2}d/w)$. The intermediate regime $\xi \sim w$ can be solved numerically, and we find that it is accurately described by a simple interpolation between the above limits with $4d/(\xi e^{-\gamma} + 2we^{-3/2})$ in the logarithm. This leads to[44]

$$g = \left[1 + \frac{2}{\pi}\frac{e^2}{\epsilon\hbar v_F}\ln\left(\frac{7.1d}{\xi + 0.8w}\right)\right]^{-1/2}. \qquad (4.1)$$

For $\epsilon = 15$, $\hbar v_F = .35$ eV nm, $\xi = 2\hbar v_F/E_{\text{gap}} \sim 30$ nm (E_{gap} is the gap of the bulk QSHI), $w = 12$ nm, and $d = 150$ nm (Ref. 45) this predicts $g \sim 0.8$. The critical exponent governing the temperature dependence of pinch-off curve (1.1) is then $\alpha_g \sim 0.02$. In the CC and II phases the conductance vanishes as T^δ with $\delta_g = g + g^{-1} - 2 \sim 0.05$.

The good news is that since g is close to 1 the low temperature scaling behavior should be accurately described by scaling function (3.9) computed in the limit $g \to 1$. The bad news is that the smallness of α_g and δ_g means that it will be difficult to see much dynamic range in the conductance as a function of temperature. Nonetheless, it may be possible to observe logarithmic corrections to the conductance as a function of temperature, and by comparing pinch-off curves at different temperatures it may be possible to observe the predicted sharpening of the transition as temperature is lowered.

The skew conductance G_{XY} is predicted be zero for noninteracting electrons and with weak interactions vanishes as T^2. This is a consequence of the unique edge state structure of the QSHI and remains robust when the interactions are weak.

To probe the critical behavior of the pinch-off transition, as well as the more exotic strong interaction phases it would be desirable to engineer structures with smaller g. Perhaps this could be accomplished by modifying either the dielectric environment or the bare Fermi velocity of the edge states. Maciejko et al.[12] suggested that this may be possible using InAs/GaSb/AlSb type-II quantum wells.[46,47]

B. Theoretical issues

Our work points to a number of theoretical problems for future study. It would be very interesting if the powerful framework of conformal field theory can be used to analyze the intermediate critical fixed point as well as the crossover scaling function. Perhaps the first place to look is $g = 1/\sqrt{3}$. Maybe it is possible to take advantage of the triangular symmetry of the QBM problem to develop a complete description of the critical fixed point, analogous to the mapping to the three-channel Kondo problem[25] and the three state Potts model[24] that apply in a different regime. In the absence of an analytic solution, this problem is amenable to a numerical Monte Carlo analysis analogous to the calculation of the resonance crossover scaling function performed in Ref. 19.

In addition, there are a number of other fixed points which we did not analyze in detail in this paper. (Recall for $g = 1 - \epsilon$ we found seven.) It would be of interest to develop a more systematic classification of all of the fixed points, analogous to the analysis of three coupled Luttinger liquids performed by Chamon and co-workers.[39,48]

APPENDIX: FOUR-TERMINAL CONDUCTANCE

The electrical response of the point contact can be characterized by a four-terminal conductance,

$$I_i = \sum_j G_{ij} V_j, \tag{A1}$$

where I_i is the current flowing into lead i and V_j is the voltage at lead j. In this appendix we will develop a convenient representation for G_{ij}. In the Appendix, Sec. I shows that G_{ij} can be characterized by a 3×3 matrix, whose entries have a clear physical meaning. This representation allows constraints due to symmetry to be expressed in a simple way, which reduces the number of independent parameters characterizing the conductance. Finally, in the Appendix, Sec. III we show how G_{ij} is related to the conductance of the SLL model computed by the Kubo formula.

1. Conductance matrix

The 4×4 matrix G_{ij} is constrained by current conservation to satisfy $\Sigma_i G_{ij} = \Sigma_j G_{ij} = 0$. In the absence of any symmetry constraints, there are thus nine independent parameters characterizing G_{ij}. In this section we will cast these nine numbers as a 3×3 matrix, in which each of the entries has a clear physical meaning. In this representation constraints due to symmetry have a simple form.

Since the four currents I_i satisfy $\Sigma_i I_i = 0$, they are determined by three *independent* currents, which we define as $I_A = (I_X, I_Y, I_Z)$ and satisfy

$$I_i = \sum_\alpha M_{iA} I_A, \tag{A2}$$

where the 4×3 matrix M_{iA} is

$$M = \frac{1}{2} \begin{pmatrix} 1 & 1 & 1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \\ 1 & -1 & -1 \end{pmatrix}. \tag{A3}$$

$I_X = I_1 + I_4$ is the total current flowing from left to right along the Hall bar, whereas $I_Y = I_1 + I_2$ is the current flowing from top to bottom. The third current $I_Z = I_1 + I_3$ is the current flowing in on opposite leads (1 and 3) and flowing out in leads 2 and 4. Similarly, the voltages V_i, which are defined up to an additive constant, define three independent voltage differences $V_\beta = (V_X, V_Y, V_Z)$, with

$$V_B = \sum_j M_{Bj}^T V_j. \tag{A4}$$

V_X biases leads 1 and 4 relative to leads 2 and 3, V_Y biases leads 1 and 2 relative to leads 3 and 4, and V_Z biases leads 1 and 3 relative to leads 2 and 4.

The new currents and voltages are then related by a 3×3 conductance matrix,

$$I_A = \sum_\beta G_{AB} V_B. \tag{A5}$$

The nine elements of G_{AB} determine the four-terminal conductance matrix,

$$G_{ij} = \sum_{AB} M_{iA} G_{AB} M_{Bj}^T. \tag{A6}$$

The elements of G_{AB} have a simple physical interpretation. G_{XX} is the "two-terminal" conductance measured horizontally in Fig. 1 by applying a voltage to leads 1 and 4 and measuring the current $I_1 + I_4$. Similarly G_{YY} is a two-terminal conductance measured vertically. G_{ZZ} describes a two-terminal conductance defined by combining the opposite leads 1 and 3 together into a single lead (and similarly for leads 2 and 4). G_{XY} is a skew conductance describing the current $I_1 + I_4$ in response to voltages applied to leads 1 and 2. The other off-diagonal conductances can be understood similarly.

2. Symmetry constraints

The form of G_{AB} simplifies considerably in the presence of symmetries.

a. Time-reversal symmetry

In the presence of time-reversal symmetry the four-terminal conductance obeys the reciprocity relation,[40] $G_{ij} = G_{ji}$. This implies $G_{AB} = G_{BA}$. Thus, with time-reversal symmetry the conductance has six independent components.

b. Spin rotational symmetry

When the spin S_z is conserved the current of up and down spins flowing into the junction must independently be conserved. It follows that

$$I_{1,\text{in}} + I_{3,\text{in}} = I_{2,\text{out}} + I_{4,\text{out}},$$

$$I_{2,\text{in}} + I_{4,\text{in}} = I_{1,\text{out}} + I_{3,\text{out}}. \tag{A7}$$

Since in the Fermi liquid lead (where the interactions have been turned off) we have $I_{i,\text{in}} = (e^2/h) V_i$, this implies that

$$I_1 + I_3 = -I_2 - I_4 = \frac{e^2}{h}(V_1 + V_3 - V_2 - V_4). \tag{A8}$$

It then follows that

$$G_{ZZ} = 2e^2/h,$$

$$G_{ZX} = G_{ZY} = 0. \tag{A9}$$

Thus, with spin conservation the conductance is characterized by three components: the two-terminal conductances G_{XX}, G_{YY} and the skew conductance G_{XY}.

The quantization of G_{ZZ} and vanishing of G_{ZB} are therefore a diagnostic for the conservation of spin. Though spin-orbit terms violating S_z conservation are generically present,

we will argue that at the low energy fixed points of physical interest the conservation of spin is restored.

c. Mirror symmetry

If the junction has a mirror symmetry under interchanging leads $(1,2) \leftrightarrow (3,4)$ or $(1,4) \leftrightarrow (2,3)$, it follows that

$$G_{XY} = 0. \tag{A10}$$

Though mirror symmetry is not generically present in a point contact we will argue that that symmetry is restored in the low energy fixed points of interest. Moreover, the *crossover* between the critical fixed point and the stable fixed point described by Eq. (1.1) is also along a line with mirror symmetry. Thus the crossover conductance is characterized by *two* parameters, G_{XX} and G_{YY}, which are simply the two-terminal conductances.

d. Critical conductance

At the transition, where the point contact is just being pinched off the two-terminal conductances must be equal,

$$G_{XX} = G_{YY} \equiv G^*. \tag{A11}$$

In addition, we will argue that this fixed point also has spin rotational symmetry and mirror symmetry. Thus, the critical four-terminal conductance G_{ij} depends on a *single* parameter G^*.

3. Relation to Kubo conductance

In this section we relate the conductance matrix G_{AB} to the conductances of the SLL model, which can be computed with the Kubo formula. There are two issues to be addressed. First is to translate G_{AB} into the spin and charge conductances of the SLL model. Second, we must relate the physical conductance measured with leads to the conductance computed with the Kubo formula. The Kubo conductance describes the response of an infinite Luttinger liquid, where the limit $L \to \infty$ is taken *before* $\omega \to 0$. This does not take into account the contact resistance between the Luttinger liquid and the electron reservoir where the voltage is defined. An appropriate model to account for this is to consider a 1D model for the leads in which the Luttinger parameter $g=1$ for $x > L$.[35,36]

In this section we assume time-reversal symmetry and that spin is conserved. In this case we may define the charge and spin currents in the Fermi liquid leads ($x > L$) to be

$$I_\rho = I_{1,\text{in}} + I_{4,\text{in}} - I_{1,\text{out}} - I_{4,\text{out}},$$

$$I_\sigma = I_{1,\text{in}} - I_{4,\text{in}} + I_{1,\text{out}} - I_{4,\text{out}}. \tag{A12}$$

Similarly, we define charge and spin voltages,

$$V_\rho = (V_1 + V_4 - V_2 - V_3)/2,$$

$$V_\sigma = (V_1 - V_4 + V_2 - V_3)/2. \tag{A13}$$

These are related by the conductance matrix

$$I_\alpha = G_{\alpha\beta} V_\beta, \tag{A14}$$

where $\alpha, \beta = \rho, \sigma$. By comparing Eqs. (A5) and (A14) it is clear that

$$G_{XX} = G_{\rho\rho},$$

$$G_{YY} = 2e^2/h - G_{\sigma\sigma},$$

$$G_{XY} = G_{\rho\sigma} = -G_{\sigma\rho}. \tag{A15}$$

$G_{\alpha\beta}$ can be computed using the Kubo formula using the model in which the interactions are turned off for $x > L$. It is useful, however, to relate this to the Kubo conductance $G^K_{\alpha\beta}$ of an infinite Luttinger liquid. This can be done by relating the voltage $V_{\alpha=\rho,\sigma}$ of the Fermi liquid leads with $g_\rho = g_\sigma = 1$ to the voltage $\bar{V}_\alpha$ of the incoming chiral modes of the Luttinger liquid with $g_\rho = g$ and $g_\sigma = 1/g$. By matching the boundary conditions at $x = L$ this contact resistance has the form

$$\tilde{V}_\alpha - V_\alpha = R^c_{\alpha\beta} I_\beta \tag{A16}$$

with

$$R^c_{\alpha\beta} = \frac{h}{e^2} \frac{g_\alpha - 1}{2g_\alpha} \delta_{\alpha\beta}. \tag{A17}$$

The Kubo formula with infinite leads relates $I_\alpha = G^K_{\alpha\beta} V_\beta$. Eliminating $\bar{V}_\alpha$ from Eqs. (A16) and (A17) gives the matrix relation[39]

$$G_{\alpha\beta} = [(I - R_c G^K)^{-1} G^K]_{\alpha\beta}. \tag{A18}$$

When there is mirror symmetry, so that $G_{XY} = \mu_{\rho\sigma} = 0$, the conductance matrix is diagonal, so that Eq. (A18) simplifies. In that case we find

$$G_{XX} = \frac{G^K_{\rho\rho}}{1 - R_{\rho\rho} G^K_{\rho\rho}},$$

$$G_{YY} = 2\frac{e^2}{h} - \frac{G^K_{\sigma\sigma}}{1 - R_{\sigma\sigma} G^K_{\sigma\sigma}}. \tag{A19}$$

---

[1] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 226801 (2005).

[2] B. A. Bernevig and S. C. Zhang, Phys. Rev. Lett. **96**, 106802 (2006).

[3] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 146802 (2005).

[4] F. D. M. Haldane, Phys. Rev. Lett. **61**, 2015 (1988).

[5] B. A. Bernevig, T. L. Hughes, and S. C. Zhang, Science **314**, 1757 (2006).

[6] M. König, S. Wiedmann, C. Brüne, A. Roth, H. Buhmann, L. Molenkamp, X. L. Qi, and S. C. Zhang, Science **318**, 766 (2007).

[7] C. Brüne, A. Roth, E. G. Novik, M. König, H. Buhmann, E. M. Hankiewicz, W. Hanke, J. Sinova, and L. W. Molenkamp, arXiv:0812.3768 (unpublished); A. Roth, C. Brüne, H. Buh-

mann, L. W. Molenkamp, J. Maciejko, X. L. Qi, and S. C. Zhang, arXiv:0905.0365 (unpublished);

[8] C. Wu, B. A. Bernevig, and S. C. Zhang, Phys. Rev. Lett. **96**, 106401 (2006).

[9] C. Xu and J. E. Moore, Phys. Rev. B **73**, 045322 (2006).

[10] C. Y. Hou, E. A. Kim, and C. Chamon, Phys. Rev. Lett. **102**, 076602 (2009).

[11] A. Ström and H. Johannesson, Phys. Rev. Lett. **102**, 096806 (2009).

[12] J. Maciejko, C. Liu, Y. Oreg, X. L. Qi, C. Wu, and S. C. Zhang, arXiv:0901.1685 (unpublished).

[13] Y. Tanaka and N. Nagaosa, arXiv:0904.1453 (unpublished).

[14] K. T. Law, C. Y. Seng, P. A. Lee, and T. K. Ng, arXiv:0904.2262 (unpublished).

[15] A. Furusaki and N. Nagaosa, Phys. Rev. B **47**, 4631 (1993).

[16] C. L. Kane and M. P. A. Fisher, Phys. Rev. B **46**, 15233 (1992).

[17] Here our definition of g_ρ and g_σ differs by a factor of 2 from Ref. 16.

[18] C. L. Kane and M. P. A. Fisher, Phys. Rev. Lett. **68**, 1220 (1992).

[19] K. Moon, H. Yi, C. L. Kane, S. M. Girvin, and M. P. A. Fisher, Phys. Rev. Lett. **71**, 4381 (1993).

[20] F. P. Milliken, C. P. Umbach, and R. A. Webb, Solid State Commun. **97**, 309 (1996).

[21] C. L. Kane and M. P. A. Fisher, Phys. Rev. B **46**, 7268 (1992).

[22] P. Fendley, A. W. W. Ludwig, and H. Saleur, Phys. Rev. Lett. **74**, 3005 (1995); P. Fendley and H. Saleur, Phys. Rev. B **54**, 10845 (1996).

[23] J. L. Cardy, Nucl. Phys. B **240**, 514 (1984).

[24] I. Affleck, M. Oshikawa, and H. Saleur, Nucl. Phys. B **594**, 535 (2001).

[25] H. Yi and C. L. Kane, Phys. Rev. B **57**, R5579 (1998); H. Yi, *ibid.* **65**, 195101 (2002).

[26] A. W. W. Ludwig and I. Affleck, Nucl. Phys. B **428**, 545 (1994).

[27] K. A. Matveev, D. Yue, and L. I. Glazman, Phys. Rev. Lett. **71**,

3351 (1993); D. Yue, L. I. Glazman, and K. A. Matveev, Phys. Rev. B **49**, 1966 (1994).

[28] A. Schmid, Phys. Rev. Lett. **51**, 1506 (1983).

[29] M. P. A. Fisher and W. Zwerger, Phys. Rev. B **32**, 6190 (1985).

[30] F. Guinea, V. Hakim, and A. Muramatsu, Phys. Rev. Lett. **54**, 263 (1985).

[31] In Ref. 16 $\widetilde{\theta}_{\sigma,\rho}$ was referred to as $\varphi_{\sigma,\rho}$. To avoid confusion with the φ_σ in Sec. II B 5 we follow the notation in Ref. 24.

[32] A. O. Caldeira and A. J. Leggett, Ann. Phys. **149**, 374 (1983).

[33] K. Itai, Phys. Rev. Lett. **58**, 602 (1987).

[34] G. T. Zimanyi, K. Vladar, and A. Zawadowski, Phys. Rev. B **36**, 3186 (1987).

[35] D. L. Maslov and M. Stone, Phys. Rev. B **52**, R5539 (1995).

[36] I. Safi and H. J. Schulz, Phys. Rev. B **52**, R17040 (1995).

[37] V. V. Ponomarenko, Phys. Rev. B **52**, R8666 (1995).

[38] A. Kawabata, J. Phys. Soc. Jpn. **65**, 30 (1996).

[39] C. Chamon, M. Oshikawa, and I. Affleck, Phys. Rev. Lett. **91**, 206403 (2003); M. Oshikawa, C. Chamon, and I. Affleck, J. Stat. Mech.: Theory Exp. (2006) 02008.

[40] M. Büttiker, Phys. Rev. Lett. **57**, 1761 (1986).

[41] K. D. Schotte and U. Schotte, Phys. Rev. **182**, 479 (1969).

[42] C. L. Kane, K. A. Matveev, and L. I. Glazman, Phys. Rev. B **49**, 2253 (1994).

[43] L. I. Glazman, I. M. Ruzin, and B. I. Shklovskii, Phys. Rev. B **45**, 8454 (1992).

[44] A similar estimate was given in Ref. 12 without the numerical factor in the logarithm.

[45] M. König, H. Buhmann, L. W. Molenkamp, T. Hughes, C.-X. Liu, X.-L. Qi, and S.-C. Zhang, J. Phys. Soc. Jpn. **77**, 031007 (2008).

[46] L. J. Cooper, N. K. Patel, V. Drouot, E. H. Linfield, D. A. Ritchie, and M. Pepper, Phys. Rev. B **57**, 11915 (1998).

[47] C. Liu, T. L. Hughes, X. L. Qi, K. Wang, and S. C. Zhang, Phys. Rev. Lett. **100**, 236601 (2008).

[48] C. Y. Hou and C. Chamon, Phys. Rev. B **77**, 155422 (2008).

P Selected for a Viewpoint in *Physics*

PHYSICAL REVIEW LETTERS

Majorana Fermions and Non-Abelian Statistics in Three Dimensions

Jeffrey C. Y. Teo and C. L. Kane

Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA
(Received 25 September 2009; published 25 January 2010)

We show that three dimensional superconductors, described within a Bogoliubov–de Gennes framework, can have zero energy bound states associated with pointlike topological defects. The Majorana fermions associated with these modes have non-Abelian exchange statistics, despite the fact that the braid group is trivial in three dimensions. This can occur because the defects are associated with an orientation that can undergo topologically nontrivial rotations. A feature of three dimensional systems is that there are "braidless" operations in which it is possible to manipulate the ground state associated with a set of defects without moving or measuring them. To illustrate these effects, we analyze specific architectures involving topological insulators and superconductors.

A fundamental feature of quantum theory is the quantum statistics obeyed by identical particles. For ordinary particles, Bose and Fermi statistics are the only possibilities. Emergent excitations in correlated many particle systems, however, can exhibit fractional [1–4] and non-Abelian [5] statistics. The simplest non-Abelian excitations, known as Ising anyons [6], are Majorana fermion states associated with zero energy modes that occur in the Bogoliubov–de Gennes (BdG) description of a paired condensate [7]. They have been predicted in a variety of two dimensional (2D) electronic systems, including the $\nu = 5/2$ quantum Hall effect [8], chiral p-wave superconductors (SCs) [9], and SC-topological insulator (TI) structures [10]. The ground state of $2N$ Ising anyons has a 2^N degeneracy, and when identical particles are exchanged the state undergoes a non-Abelian unitary transformation [11,12]. Recent interest in non-Abelian statistics has been heightened by the proposal to use these features for topological quantum computation [13].

Fractional and non-Abelian statistics are usually associated with 2D because in 3D, performing an exchange twice is topologically trivial. In this Letter, we show that in 3D, Majorana fermion states are associated with pointlike topological defects, and that they obey non-Abelian exchange statistics, despite the triviality of braids. Our motivation came from the study of 3D SC-TI structures, where Majorana fermions arise in a variety of ways, such as (i) vortices at SC-TI interfaces [10], (ii) SC-magnet interfaces at the edge of a 2D TI [14–16], and (iii) band inversion domain walls along a SC vortex line. While the Majorana fermions in these cases can be identified using 1D or 2D effective theories, they must occur in a more general 3D theory. To unify them, we introduce a $\mathbb{Z}_2$ topological index that locates the zero modes in a generic 3D BdG theory. We then study a minimal 8-band model in which the defects can be understood as hedgehogs in a three component vector field. Ising non-Abelian exchange statistics arise because the hedgehogs have an *orientation* that can undergo nontrivial rotations. We will illustrate the

intrinsic three dimensionality of the Majorana states by considering specific architectures involving SCs and TIs. A feature in 3D is the existence of "braidless" operations, in which the quantum information encoded in the Majorana states can be manipulated without moving or measuring [17] them.

To determine whether a Majorana mode is enclosed in a volume V, we topologically classify BdG Hamiltonians on ∂V, the 2D surface V. We assume the Hamiltonian varies slowly, so we can consider adiabatic changes as a function of two parameters $\mathbf{r}$ characterizing ∂V. The problem is then to classify particle-hole (PH) symmetric BdG Hamiltonians $\mathcal{H}(\mathbf{k}, \mathbf{r})$, where $\mathbf{k}$ is defined in a 3D Brillioun zone (a torus T^3) and $\mathbf{r}$ is defined on a 2-sphere S^2. PH symmetry is defined by an antiunitary operator Ξ satisfying $\Xi^2 = 1$ and $\mathcal{H}(\mathbf{k}, \mathbf{r}) = -\Xi\mathcal{H}(-\mathbf{k}, \mathbf{r})\Xi^{-1}$. Assuming no other symmetries, this corresponds to class D of the general scheme [18,19]. Since V may or may not enclose a zero mode, we expect a $\mathbb{Z}_2$ classification—a fact that can be established using methods of K theory.

A formula for the topological invariant can be derived using a method similar to Qi, Hughes, and Zhang's [20] formulation of the invariant characterizing a 3D strong TI [21]. We introduce a one parameter deformation $\tilde{\mathcal{H}}(\lambda, \mathbf{k}, \mathbf{r})$ that adiabatically connects $\mathcal{H}(\mathbf{k}, \mathbf{r})$ at $\lambda = 0$ to a trivial Hamiltonian independent of $\mathbf{k}$ and $\mathbf{r}$ at $\lambda = 1$, while violating PH symmetry. PH symmetry can then be restored by including a mirror image $\tilde{\mathcal{H}}(\lambda, \mathbf{k}, \mathbf{r}) = -\Xi\tilde{\mathcal{H}}(-\lambda, -\mathbf{k}, \mathbf{r})\Xi^{-1}$ for $-1 < \lambda < 0$. For $\lambda = \pm 1$, $\mathbf{k}$, $\mathbf{r}$ can be replaced by a single point, so the 6 parameter space $(\lambda, \mathbf{k}, \mathbf{r}) \sim \Sigma(T^3 \times S^2)$ (Σ denotes the suspension) has no boundary. $\tilde{\mathcal{H}}$ defined on this space is characterized by its integer valued third Chern character [22],

$$\mathrm{Ch}_3[\mathcal{F}] = \frac{1}{3!}\left(\frac{i}{2\pi}\right)^3 \int_{\Sigma(T^3 \times S^2)} \mathrm{Tr}[\mathcal{F} \wedge \mathcal{F} \wedge \mathcal{F}]. \quad (1)$$

Here, $\mathcal{F} = d\mathcal{A} + \mathcal{A} \wedge \mathcal{A}$ follows from the non-Abelian Berry's connection $\mathcal{A}_{ij} = \langle u_i | d | u_j \rangle$ associated with the

negative energy eigenstates of $\tilde{\mathcal{H}}$. Because of PH symmetry, the contributions to (1) for $\lambda > 0$ and $\lambda < 0$ are equal. Moreover, since it is a total derivative, $\text{Tr}[\mathcal{F}^3] = d\mathcal{Q}_5$ (omitting the $\wedge$'s), where the Chern-Simons 5 form is [22]

$$\mathcal{Q}_5 = \text{Tr}[\mathcal{A} \wedge (d\mathcal{A})^2 + (3/2)\mathcal{A}^3 \wedge d\mathcal{A} + (3/5)\mathcal{A}^5]. \quad (2)$$

The integral over $\lambda > 0$ can be then be pushed to the boundary $\lambda = 0$ so that

$$\text{Ch}_3[\mathcal{F}] = \frac{2}{3!}\left(\frac{i}{2\pi}\right)^3 \int_{T^3 \times S^2} \mathcal{Q}_5. \quad (3)$$

A different deformation $\tilde{\mathcal{H}}$ can change $\text{Ch}_3[\mathcal{F}]$, but PH symmetry requires the change is an *even* integer. Likewise, the right hand side of (3) is only gauge invariant up to an even integer. The parity of (3) defines a $\mathbb{Z}_2$ topological invariant. We write it as

$$\mu = \int_{S^2} \boldsymbol{\omega}(\mathbf{r}) \cdot d\mathbf{A} \qquad \text{mod } 2 \quad (4)$$

where the gauge dependent Chern-Simons flux $\boldsymbol{\omega} = (\omega^1, \omega^2, \omega^3)$ is defined by integrating out $\mathbf{k}$,

$$\frac{1}{2}\epsilon_{ijk}\omega^i(\mathbf{r})dx^j \wedge dx^k = \frac{1}{3}\left(\frac{i}{2\pi}\right)^3 \int_{T^3} \mathcal{Q}_5. \quad (5)$$

It is natural to associate μ with the presence of a zero mode—a fact that will be checked explicitly below.

We now introduce a minimal model that leads to an appealing physical interpretation for $\boldsymbol{\omega}(\mathbf{r})$. Since μ is based on $\text{Ch}_3[\mathcal{F}]$, we expect a minimum of 8 bands is required. Consider a model parameterized by a three component vector field $\mathbf{n}$ of the form,

$$\mathcal{H} = -i\gamma_a\partial_a + \Gamma_a n_a(\mathbf{r}). \quad (6)$$

Here, γ_a and Γ_a ($a = 1, 2, 3$) are 8×8 Dirac matrices satisfying $\{\Gamma_a, \Gamma_b\} = \{\gamma_a, \gamma_b\} = 2\delta_{ab}$ and $\{\Gamma_a, \gamma_b\} = 0$. $\mathcal{H}$ respects PH symmetry provided $\Xi\Gamma_a\Xi^{-1} = -\Gamma_a$ and $\Xi\gamma_a\Xi^{-1} = \gamma_a$. For $\mathbf{n}(\mathbf{r}) = \mathbf{n}_0$, $\mathcal{H}$ has eigenvalues $E(\mathbf{k}) = \pm(|\mathbf{k}|^2 + |\mathbf{n}_0|^2)^{1/2}$ so that for $\mathbf{n}_0 \neq 0$, there is a gap $2|\mathbf{n}_0|$. The seventh Dirac matrix $\gamma_5 \equiv i\prod_a \gamma_a\Gamma_a$ is not an allowed mass term because $\Xi\gamma_5\Xi^{-1} = \gamma_5$. A more general Hamiltonian could also involve products of the Dirac matrices, but such Hamiltonians can be homotopically deformed to the form of (6) without closing the gap [19]. To regularize (6) at $|\mathbf{k}| \to \infty$, we include an additional term $\epsilon|\mathbf{k}|^2\Gamma_3$, so $\mathbf{k}$ can be defined on a compact Brillouin zone S^3. The analysis is simplest for $\epsilon \to 0$, where the low energy properties are isotropic in $\mathbf{n}$.

$\mathcal{H}$ can be physically motivated by considering a BdG Hamiltonian describing ordinary and topological insulators coexisting with superconductivity. The Dirac matrices are specified by three sets of Pauli matrices: $\vec{\tau}$ for PH space, $\vec{\sigma}$ for spin, and $\vec{\mu}$ for an orbital degree of freedom. We identify $\vec{\gamma} = \tau_z\mu_z\vec{\sigma}$, $\Gamma_1 = \tau_x$, $\Gamma_2 = \tau_y$, and $\Gamma_3 = \tau_z\mu_x$,

along with $\Xi = \sigma_y\tau_yK$. $\mathbf{n}$ is then (Δ_1, Δ_2, m), where $\Delta = \Delta_1 + i\Delta_2$ is a SC order parameter and m is a mass describing a band inversion. For $\Delta = 0$, (6) is a doubled version of the model for a 3D TI discussed in Ref. [20]. For $\epsilon > 0$, $m > 0$ describes a trivial insulator, while $m < 0$ describes a TI with a band inversion near $\mathbf{k} = 0$. An interface where m changes sign corresponds to the surface of a TI, which has gapless surface states. Introducing $\Delta \neq 0$ to the interface then describes the proximity induced SC state [10].

To locate the zero modes, we take $\mathbf{n}(\mathbf{r})$ to vary slowly with $\mathbf{r}$ and evaluate $\boldsymbol{\omega}(\mathbf{r})$ using (1)–(5). This can be done by noting that $\mathcal{H}$ defines a 6 component unit vector given by the direction $\hat{\mathbf{d}}^5(\mathbf{k}, \mathbf{r})$ of $(k_1, k_2, k_3, n_1 + \epsilon|\mathbf{k}|^2, n_2, n_3)$ on S^5. The deformed Hamiltonian $\tilde{\mathcal{H}}$ can be defined by adding $\lambda\gamma_5$ to $\mathcal{H}$ so that $\tilde{\mathcal{H}}$ defines a vector on S^6 given by $\hat{\mathbf{d}}^6 = (\sqrt{1 - \lambda^2}\hat{\mathbf{d}}^5, \lambda)$. $\text{Ch}_3[\mathcal{F}]$ is the volume on S^6 swept out by $\hat{\mathbf{d}}^6(\lambda, \mathbf{k}, \mathbf{r})$. Then, $\int \mathcal{Q}_5$ is the volume in the "northern hemisphere" of S^6 swept out by $\hat{\mathbf{d}}^5(\mathbf{k}, \mathbf{r})$, which is confined to the "equator," $\lambda = 0$. This is then related to the area on S^5 swept out by $\hat{\mathbf{d}}^5(\mathbf{k}, \mathbf{r})$. Performing the integral on $\mathbf{k}$ for $\epsilon \to 0$ gives

$$\omega^i(\mathbf{r}) = \frac{1}{8\pi}\epsilon^{ijk}\hat{\mathbf{n}} \cdot \partial_j\hat{\mathbf{n}} \times \partial_k\hat{\mathbf{n}}, \quad (7)$$

where $\hat{\mathbf{n}} = \mathbf{n}/|\mathbf{n}|$. Thus, the topological charge that signals a zero mode inside V is the parity of the S^2 winding number of $\hat{\mathbf{n}}$ on ∂V. Zero modes are associated with *hedgehogs* in $\hat{\mathbf{n}}(\mathbf{r})$. A simple example of a hedgehog is an SC vortex at the interface between a TI and an insulator. Though the hedgehog topological charge can be any integer, an even integer in (4) can be unwound by a $\mathbf{k}$ and $\mathbf{r}$ dependent gauge transformation.

The presence of a zero mode associated with a hedgehog can be demonstrated with a simple linear model $n_a(\mathbf{r}) = M_{ab}r_b$, which has a hedgehog with charge $\text{sgn}(\det[M])$ at $\mathbf{r} = 0$. This is solved by expressing M in terms of its principle axes: $M = \mathcal{O}_1^T\text{diag}(M_1, M_2, M_3)\mathcal{O}_2$, where $\mathcal{O}_1$ and $\mathcal{O}_2$ are orthogonal matrices that diagonalize MM^T and M^TM, respectively. Defining $r_a' = \mathcal{O}_{1ab}r_b$, $n_a' = \mathcal{O}_{2ab}n_b$, $\gamma_a' = \mathcal{O}_{1ab}^T\gamma_b$, and $\Gamma_a' = \mathcal{O}_{2ab}^T\Gamma_b$, it is straightforward to express $\mathcal{H}^2$ as three independent harmonic oscillators,

$$\mathcal{H}^2 = \sum_a M_a(2n_a + 1 - \xi_a) \quad (8)$$

where n_a are oscillator quantum numbers and $\xi_a = i\gamma_a'\Gamma_a'$ are commuting operators. There is a single zero energy state with $n_a = 0$ and $\xi_a = 1$. This zero mode is the nondegenerate eigenstate with eigenvalue 3 of $\sum_a \xi_a = i\sum_{a,b}\gamma_a\mathcal{O}_{ab}\Gamma_b$, where $\mathcal{O} = \mathcal{O}_1\mathcal{O}_2^T$.

A key feature of the zero mode is its dependence on the relative *orientation* $\mathcal{O}$ of the principle axes in $\mathbf{r}$ and $\mathbf{n}$ space. This can lead to a non trivial holonomy when M (and hence $\mathcal{O}$) varies. We construct the zero mode by starting with $|\Psi_0\rangle$ which satisfies $\sum_a i\gamma_a\Gamma_a|\Psi_0\rangle = 3|\Psi_0\rangle$

and then doing a unitary transformation that takes Γ_a to $\mathcal{O}_{ab}\Gamma_b$. If we parameterize the rotation $\mathcal{O}$ with a vector $\mathbf{\Omega}$ specifying the axis and angle $|\mathbf{\Omega}| \leq \pi$, then

$$|\Psi(\mathbf{\Omega})\rangle = e^{\epsilon_{abc}\Gamma_a\Gamma_b\Omega_c/4}|\Psi_0\rangle. \qquad (9)$$

This gauge satisfies $\langle\Psi(\mathbf{\Omega})|d\Psi(\mathbf{\Omega})\rangle = 0$ for $|\mathbf{\Omega}| < \pi$ but is not globally defined because $|\Psi(\mathbf{\Omega})\rangle = -|\Psi(-\mathbf{\Omega})\rangle$ when $|\mathbf{\Omega}| = \pi$. This reflects the nontrivial topology of $SO(3)$, characterized by the homotopy $\pi_1[SO(3)] = \mathbb{Z}_2$. When $\mathcal{O}$ varies along a nontrivial loop in $SO(3)$, the wave function of the zero mode changes sign. The associated Majorana operator γ_i (not to be confused with the Dirac matrix γ_a) also changes sign. This is one of our central results, and it is this fact that allows Ising non-Abelian statistics in 3D. A simple example of a nontrivial loop is the 2π rotation that occurs when the SC phase advances by 2π.

Though we derived it with the linear 8 band model, our conclusion that there is nontrivial holonomy for defects is more general. A formulation based on (4) will appear elsewhere. In the 8-band model, a general defect history is characterized by $\mathbf{n}(\mathbf{r}, t)$, where $\mathbf{r}$ is on a surrounding surface S^2 and t varies on a closed path S^1. These are classified by the homotopy of maps $S^2 \times S^1 \to S^2$, which were first analyzed by Pontrjagin [23], and have appeared in other physical contexts [24]. When the hedgehog's topological charge is $\pm p$, the classification is $\mathbb{Z}_{2p}$. This is related to the integer Hopf invariant for maps $S^3 \to S^2$ which applies when $p = 0$. Like the Hopf invariant, it can be understood in terms of the linking of curves in $S^2 \times S^1$ [25]. The discussion below resembles the analysis of Wilczek and Zee [2] of the statistics of Skyrmions in the $(2 + 1)D$ nonlinear σ model with a Hopf term.

To study the exchange statistics of the Majorana modes, we consider the adiabatic evolution of the state when they are exchanged. We thus consider a $3 + 1D$ history $\hat{\mathbf{n}}(\mathbf{r}, t)$ satisfying $\hat{\mathbf{n}}(\mathbf{r}, T) = \hat{\mathbf{n}}(\mathbf{r}, 0)$ with hedgehogs at $\{\mathbf{r}_i(t)\}$ with $\mathbf{r}_{1(2)}(T) = \mathbf{r}_{2(1)}(0)$. To visualize $\hat{\mathbf{n}}(\mathbf{r}, t)$, it is useful to consider the inverse image paths in $\mathbf{r}$ space that map to two specific points on S^2. Such paths begin and end on hedgehogs, and a crucial role will be played by their linking properties. Figure 1 depicts four hedgehogs, where the top two (positive) hedgehogs are interchanged. At the first step in (b), the locations of the hedgehogs have been interchanged. Since the inverse image paths have been "dragged," $\hat{\mathbf{n}}(\mathbf{r})$ is not the same as its original configuration. Panels (c)–(f) show a sequence of smooth deformations that untangle $\hat{\mathbf{n}}(\mathbf{r})$. The key point is that the two paths (which map to different points on S^2) can never cross each other. However, a path can cross itself and "reconnect," as in (c), (d), (e). The deformations from (a)–(e) preserve the orientation of the hedgehogs, but leave behind a twist. To return $\hat{\mathbf{n}}(\mathbf{r})$ to its original configuration in (f) requires a 2π rotation of one of the hedgehogs. This results in an interchange rule for the Majoranas,
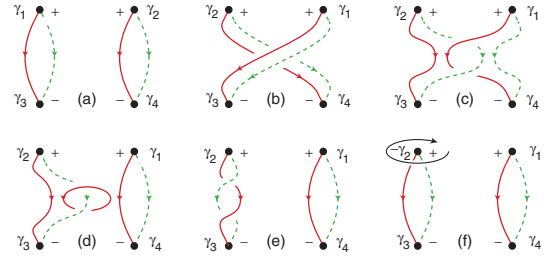
FIG. 1 (color online). Inverse image paths depicting the interchange of two hedgehogs, as described in the text. (a)–(e) show a sequence of continuous deformations of $\mathbf{n}(\mathbf{r})$ that preserve the orientation of the hedgehogs. In (f), a 2π rotation is required to return to the original configuration.

$$T_{12}: \gamma_1 \to \gamma_2; \qquad \gamma_2 \to -\gamma_1 \qquad (10)$$

analogous to the rules [11,12] for braiding vortices in 2D and can be represented by $T_{12} = \exp[\pi\gamma_1\gamma_2/4]$. Note that in (d), (e), the twist could have been left on the other side, which would have led to $T_{21} = T_{12}^{\dagger}$. The two choices for T_{12} correspond to *physically distinct* interchange trajectories that generalize the right- and left-handed braiding operations in 2D.

Performing the same interchange twice leads to a nontrivial operation, since $T_{ij}^2 = \gamma_i\gamma_j$ changes the sign of both γ_i and γ_j. This is natural in 2D because it is a noncontractable braid. In 3D, however, T_{ij}^2 can be smoothly deformed into an operation in which all particles are held fixed. Thus, there is an operation, specified by a history $\hat{\mathbf{n}}(\mathbf{r}, t)$, that rotates any pair of *stationary* hedgehogs by 2π, and implements the operation $\gamma_i\gamma_j$. The existence of such "braidless" operations is a feature of Ising non-Abelian statistics in 3D. Although these operations form an admittedly limited Abelian subgroup, they nonetheless offer a method for manipulating the quantum information encoded in the Majorana fermions without moving or measuring them.

We now illustrate these effects using specific architectures involving TIs and SCs. It is easiest to engineer Majorana modes using structures involving interfaces or vortex lines, where $\boldsymbol{\omega}$ is confined to lines or planes. Nonetheless, such structures can exhibit intrinsically 3D effects. Consider first the structure in Fig. 2(a), which involves two disconnected spherical TIs surrounded by a SC and connected to each other by a Josephson junction. Suppose that each sphere has a single $\pm$ pair of vortices so that there are 4 Majorana states on the spheres. The internal state of the Majorana fermions can be represented in a basis of eigenstates of $n = i\gamma_1\gamma_2$ and $n' = i\gamma_3\gamma_4$. For an isolated system, the parity of $n + n'$ is fixed so the system is a single qubit with basis vectors $|nn' = 00, 11\rangle$. The state can be initialized and measured in this basis with a probe that couples to both γ_1 and γ_2. Suppose γ_3 is adiabatically transported around γ_1 as shown. This is similar to a 2D braid, and it implements T_{13}^2, which interchanges $|00\rangle$ and
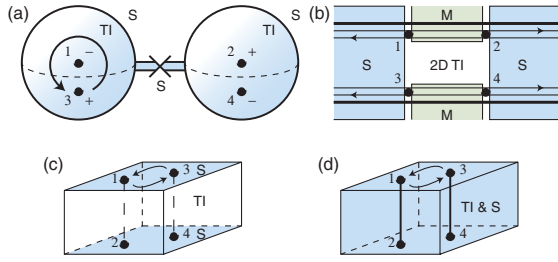
FIG. 2 (color online). Architectures demonstrating 3D Majorana states using SCs and TIs. (a) Braiding 3 around 1 can be deformed into a braidless operation. (b) A geometry for implementing braidless operations with a Josephson junction device [15]. (c), (d) Thin film geometries for interchanging and measuring Majorana states. (c) shows a layered SC-TI-SC structure, while (d) shows a thin film of a SC weakly doped TI.

$|11\rangle$. Note, however, that this "braid" can be smoothly contracted to zero by sliding the path around the other side of the sphere. In the process, however, the path crosses the junction connecting the spheres, resulting in a 2π phase slip. Thus, the braiding operation can be smoothly deformed into a "braidless" operation, where γ_1, γ_3 are held fixed, but the phase difference between the two SCs advances by 2π.

A similar but more feasible version of the braidless operation occurs for the Josephson junction structure in Fig. 2(b), which involves Majorana modes at a SC-magnet interface at the edge of a 2D TI. As argued in Ref. [15], when the phase difference across the junction is advanced by 2π, the fermion parity associated with $i\gamma_1\gamma_2$ changes, resulting in a fractional Josephson effect.

Figure 2(c) shows a TI coated on the top and bottom with SC films. In a magnetic field, Majorana states occur at vortices on both the top and bottom. If the SC films are thinner than the penetration depth, the field is constant, so the top and bottom vortices are independent. If the TI is thin, there will be a weak vertical coupling that splits nearby Majorana modes according to $n = i\gamma_1\gamma_2$ and $n' = i\gamma_3\gamma_4$. The states $|nn'\rangle$ will have slightly different charges, which may allow n, n' to be measured with a sensitive charge detector. Suppose the state is initially $|00\rangle$. Interchanging γ_1, γ_3 on the top keeping the bottom fixed leads to the entangled state $(|00\rangle + |11\rangle)/\sqrt{2}$. A variant on this geometry [Fig. 2(d)] is a thin film of a bulk SC *weakly doped* TI, in which the surface states acquire SC similar to the proximity induced state. In this case, the interchange of γ_1, γ_3 involves a reconnection of the vortex lines (which in principle have a finite energy gap) connecting them.

Recently, SC has been observed in $Cu_xBi_2Se_3$ for $x \sim 0.15$ [26]. It will be interesting to determine whether this material is in the weakly doped regime, with Majorana modes at the *ends* of vortex lines, or a more conventional

SC, which could be used in Fig. 2(c). There are certainly technical challenges associated with manipulating the vortices and measuring their charge state. Nonetheless, we hope that the prospect of detecting 3D non-Abelian statistics in such a system will provide motivation for further exploration.

[1] J. M. Leinaas and J. Myrheim, Nuovo Cimento Soc. Ital. Fis. B **37**, 1 (1977).
[2] F. Wilczek and A. Zee, Phys. Rev. Lett. **51**, 2250 (1983).
[3] B. I. Halperin, Phys. Rev. Lett. **52**, 1583 (1984).
[4] D. Arovas, J. R. Schrieffer, and F. Wilczek, Phys. Rev. Lett. **53**, 722 (1984).
[5] G. Moore and N. Read, Nucl. Phys. **B360**, 362 (1991).
[6] C. Nayak *et al.*, Rev. Mod. Phys. **80**, 1083 (2008).
[7] N. Read and D. Green, Phys. Rev. B **61**, 10267 (2000).
[8] M. Greiter, X. G. Wen, and F. Wilczek, Nucl. Phys. **B374**, 567 (1992).
[9] S. Das Sarma, C. Nayak, and S. Tewari, Phys. Rev. B **73**, 220502(R) (2006).
[10] L. Fu and C. L. Kane, Phys. Rev. Lett. **100**, 096407 (2008).
[11] C. Nayak and F. Wilczek, Nucl. Phys. **B479**, 529 (1996).
[12] D. A. Ivanov, Phys. Rev. Lett. **86**, 268 (2001).
[13] A. Kitaev, Ann. Phys. (N.Y.) **303**, 2 (2003).
[14] A. Kitaev, arXiv:cond-mat/0010440.
[15] L. Fu and C. L. Kane, Phys. Rev. B **79**, 161408(R) (2009).
[16] J. Nilsson, A. R. Akhmerov, and C. W. J. Beenakker, Phys. Rev. Lett. **101**, 120403 (2008).
[17] P. Bonderson, M. Freedman, and C. Nayak, Phys. Rev. Lett. **101**, 010501 (2008).
[18] A. P. Schnyder, S. Ryu, A. Furusaki, and A. W. W. Ludwig, Phys. Rev. B **78**, 195125 (2008); AIP Conf. Proc. **1134**, 10 (2009).
[19] A. Kitaev, AIP Conf. Proc. **1134**, 22 (2009); arXiv:0901.2686.
[20] X. L. Qi, T. L. Hughes, and S. C. Zhang, Phys. Rev. B **78**, 195424 (2008).
[21] L. Fu, C. L. Kane, and E. J. Mele, Phys. Rev. Lett. **98**, 106803 (2007); J. E. Moore and L. Balents, Phys. Rev. B **75**, 121306(R) (2007).
[22] M. Nakahara, *Geometry, Topology and Physics* (Adam Hilger, Bristol, 1990).
[23] L. S. Pontrjagin, Rec. Math. [Mat. Sbornik] N.S. **9**, 331 (1941); http://mi.mathnet.ru/eng/msb6073.
[24] J. Jäykkä and J. Hietarinta, Phys. Rev. D **79**, 125027 (2009).
[25] L. Kapitanski, in London Mathematical Society Durham Symposium (unpublished); http://www. maths.dur.ac.uk/ events/Meetings/LMS/200/OTSA/.
[26] Y. S. Hor *et al.*, Phys. Rev. Lett. (to be published).

ॐ

Topological defects and gapless modes in insulators and superconductors

Jeffrey C. Y. Teo and C. L. Kane

Department of Physics and Astronomy, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

We develop a unified framework to classify topological defects in insulators and superconductors described by spatially modulated Bloch and Bogoliubov de Gennes Hamiltonians. We consider Hamiltonians $\mathcal{H}(\mathbf{k}, \mathbf{r})$ that vary slowly with adiabatic parameters $\mathbf{r}$ surrounding the defect and belong to any of the ten symmetry classes defined by time-reversal symmetry and particle-hole symmetry. The topological classes for such defects are identified and explicit formulas for the topological invariants are presented. We introduce a generalization of the bulk-boundary correspondence that relates the topological classes to defect Hamiltonians to the presence of protected gapless modes at the defect. Many examples of line and point defects in three-dimensional systems will be discussed. These can host one dimensional chiral Dirac fermions, helical Dirac fermions, chiral Majorana fermions, and helical Majorana fermions, as well as zero-dimensional chiral and Majorana zero modes. This approach can also be used to classify temporal pumping cycles, such as the Thouless charge pump, as well as a fermion parity pump, which is related to the Ising non-Abelian statistics of defects that support Majorana zero modes.

I. INTRODUCTION

The classification of electronic phases according to topological invariants is a powerful tool for understanding and predicting the behavior of matter. This approach was pioneered by Thouless, *et al.*[1] (TKNN), who identified the integer topological invariant characterizing the two-dimensional (2D) integer quantum-Hall state. The TKNN invariant n gives the Hall conductivity $\sigma_{xy} = ne^2/h$ and characterizes the Bloch Hamiltonian $\mathcal{H}(\mathbf{k})$, defined as a function of $\mathbf{k}$ in the magnetic Brillouin zone. It may be expressed as the first Chern number associated with the Bloch wave functions of the occupied states. A fundamental consequence of this topological classification is the *bulk-boundary correspondence*, which relates the topological class of the bulk system to the number of gapless chiral fermion edge states on the sample boundary.

Recent interest in topological states[2–4] has been stimulated by the realization that the combination of time-reversal symmetry and the spin-orbit interaction can lead to topological insulating electronic phases[5–10] and by the prediction[11–13] and observation[14–26] of these phases in real materials. A topological insulator is a two- or three-dimensional material with a bulk energy gap that has gapless modes on the edge or surface that are protected by time-reversal symmetry. The bulk boundary correspondence relates these modes to a $\mathbb{Z}_2$ topological invariant characterizing time-reversal invariant Bloch Hamiltonians. Signatures of these protected boundary modes have been observed in transport experiments on 2D HgCdTe quantum wells[14–16] and in photoemission and scanning tunnel microscope experiments on three-dimensional (3D) crystals of $Bi_{1-x}Sb_x$,[17–19] Bi_2Se_3,[20] Bi_2Te_3,[22,23,25] and Sb_2Te_3.[26] Topological insulator behavior has also been predicted in other classes of materials with strong spin-orbit interactions.[27–33]

Superconductors, described within a Bogoliubov de Gennes (BdG) framework can similarly be classified topologically.[34–37] The Bloch-BdG Hamiltonian $\mathcal{H}_{BdG}(\mathbf{k})$ has a structure similar to an ordinary Bloch Hamiltonian, except that it has an exact particle-hole symmetry that reflects the particle-hole redundancy inherent to the BdG theory. Topological superconductors are also characterized by gapless boundary modes. However, due to the particle-hole redundancy, the boundary excitations are Majorana fermions. The simplest model topological superconductor is a weakly paired spinless p wave superconductor in one-dimensional (1D),[38] which has zero-energy Majorana bound states at its ends. In 2D, a weakly paired $p_x + ip_y$ superconductor has a chiral Majorana edge state.[39] Sr_2RuO_4 is believed to exhibit a triplet $p_x + ip_y$ state.[40] The spin degeneracy, however, leads to a doubling of the Majorana edge states. Though undoubled topological superconductors remain to be discovered experimentally, superfluid ^{3}He B is a related topological phase[34,35,37,41,42] and is predicted to exhibit 2D gapless Majorana modes on its surface. Related ideas have also been used to topologically classify Fermi surfaces.[43]

Topological insulators and superconductors fit together into an elegant mathematical framework that generalizes the above classifications.[35,36] The topological classification of a general Bloch or BdG theory is specified by the dimension d and the ten Altland-Zirnbauer symmetry classes[44] characterizing the presence or absence of particle-hole, time-reversal, and/or chiral symmetry. The topological classifications, given by $\mathbb{Z}$, $\mathbb{Z}_2$, or 0 show a regular pattern as a function of symmetry class and d, and can be arranged into a *periodic table* of topological insulators and superconductors. Each nontrivial entry in the table is predicted, via the bulk-boundary correspondence, to have gapless boundary states.

Topologically protected zero modes and gapless states can also occur at topological defects, and have deep implications in both field theory and condensed matter physics.[41,45–47] A simple example is the zero-energy Majorana mode that occurs at a vortex in a $p_x + ip_y$ superconductor.[39] Similar Majorana bound states can be engineered using three dimensional heterostructures that combine ordinary superconductors and topological insulators,[48] as well as semiconductor structures

FIG. 1. (Color online) Topological defects characterized by a D parameter family of d-dimensional Bloch-BdG Hamiltonians. Line defects correspond to $d-D=2$ while point defects correspond to $d-D=1$. Temporal cycles for point defects correspond to $d-D=0$.

that combine superconductivity, magnetism, and strong spin-orbit interactions.[49–52] Recently, we showed that the existence of a Majorana bound state at a point defect in a three dimensional Bogoliubov de Gennes theory is related to a $\mathbb{Z}_2$ topological invariant that characterizes a family of Bogoliubov de Gennes Hamiltonians $\mathcal{H}_{\mathrm{BdG}}(\mathbf{k},\mathbf{r})$ defined for $\mathbf{r}$ on a surface surrounding the defect.[53] This suggests that a more general formulation of topological defects and their corresponding gapless modes should be possible.

In this paper we develop a general theory of topological defects and their associated gapless modes in Bloch and Bloch-BdG theories in all symmetry classes. As in Ref. 53, we assume that far away from the defect the Hamiltonian varies slowly in real space, allowing us to consider adiabatic changes in the Hamiltonian as a function of the real space position $\mathbf{r}$. We thus seek to classify Hamiltonians $\mathcal{H}(\mathbf{k},\mathbf{r})$, where $\mathbf{k}$ is defined in a d-dimensional Brillouin zone (a torus T^d), and $\mathbf{r}$ is defined on a D-dimensional surface S^D surrounding the defect. A similar approach can be used to classify cyclic temporal variations in the Hamiltonian, which define adiabatic pumping cycles. Hereafter we will drop the BdG subscript on the Hamiltonian with the understanding that the symmetry class dictates whether it is a Bloch or BdG Hamiltonian.

In Fig. 1 we illustrate the types of topological defects that can occur in $d=1$, 2, or 3. For $D=0$ we regard S^0 as two points ($\{-1,+1\}$). Our topological classification then classifies the *difference* of $\mathcal{H}(\mathbf{k},+1)$ and $\mathcal{H}(\mathbf{k},-1)$. A nontrivial difference corresponds to an interface between two topologically distinct phases. For $D=1$ the one parameter families of Hamiltonians describe line defects in $d=3$ and point defects in $d=2$. For $D=1$ it could correspond to an adiabatic temporal cycle $H(\mathbf{k},t)$. Similarly for $D=2$, the two parameter family describes a point defect for $d=3$ or an adiabatic cycle for a point defects in $d=2$.

Classifying the D parameter families of d-dimensional Bloch-BdG Hamiltonians subject to symmetries leads to a generalization of the periodic table discussed above. The original table corresponds to $D=0$. For $D>0$ we find that for a given symmetry class the topological classification ($\mathbb{Z}$, $\mathbb{Z}_2$, or 0) depends only on

$$\delta = d - D. \tag{1.1}$$

Thus, all line defects with $\delta=2$ have the same topological classification, irrespective of d, as do point defects with $\delta=1$ and pumping cycles with $\delta=0$. Though the classifications depend only on δ, the *formulas* for the topological invariants depend on both d and D.

This topological classification of $\mathcal{H}(\mathbf{k},\mathbf{r})$ suggests a generalization of the bulk-boundary correspondence that relates the topological class of the Hamiltonian characterizing the defect to the structure of the protected modes associated with the defect. This has a structure reminiscent of a mathematical *index theorem*[54] that relates a topological index to an analytical index that counts the number of zero modes.[41,45,46,55–59] In this paper we will not attempt to *prove* the index theorem. Rather, we will observe that the topological classes for $\mathcal{H}(\mathbf{k},\mathbf{r})$ coincide with the expected classes of gapless defect modes. In this regards the dependence of the classification on δ in Eq. (1.1) is to be expected. For example, a point defect at the end of a one-dimensional system ($\delta=1-0$) has the same classification as a point defects in two dimensions ($\delta=2-1$) and three dimensions ($\delta=3-2$).

We will begin in Sec. II by describing the generalized periodic table. We will start with a review of the Altland Zirnbauer symmetry classes[44] and a summary of the properties of the table. In Appendix A we will justify this generalization of the table by introducing a set of mathematical mappings that relate Hamiltonians in different dimensions and different symmetry classes. In addition to establishing that the classifications depend only on $\delta=d-D$, these mappings allow other features of the table, already present for $D=0$ to be easily understood, such as the pattern in which the classifications vary as a function of symmetry class as well as the Bott periodicity of the classes as a function of d.

In Secs. III and IV we will outline the physical consequences of this theory by discussing a number of examples of line and point defects in different symmetry classes and dimensions. The simplest example is that of a line defect in a 3D system with no symmetries. In Sec. III A we will show that the presence of a 1D *chiral Dirac fermion* mode (analogous to an integer quantum-Hall edge state) on the defect is associated with an integer topological invariant that may be interpreted as the winding number of the "θ" term that characterizes the magnetoelectric polarizability.[10] This description unifies a number of methods for "engineering" chiral Dirac fermions, which will be described in several illustrative examples.

Related topological invariants and illustrative examples will be presented in Secs. III B–III E for line defects in other symmetry classes that are associated with gapless 1D helical Dirac fermions, 1D chiral Majorana fermions, and 1D helical Majorana fermions. In Sec. IV we will consider point defects in 1D models with chiral symmetry such as the Jackiw-Rebbi model[45] or the Su, Schrieffer, Heeger model,[47] and in superconductors without chiral symmetry that exhibit Majorana bound states or Majorana doublets. These will also be related to the early work of Jackiw and Rossi[46] on Majorana modes at point defects in a model with chiral symmetry.

TABLE I. Periodic table for the classification of topological defects in insulators and superconductors. The rows correspond to the different Altland Zirnbauer (AZ) symmetry classes while the columns distinguish different dimensionalities, which depend only on $\delta=d-D$.

		Symmetry			$\delta=d-D$							
s	AZ	Θ^2	Ξ^2	Π^2	0	1	2	3	4	5	6	7
0	A	0	0	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0
1	AIII	0	0	1	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$	0	$\mathbb{Z}$
0	AI	1	0	0	$\mathbb{Z}$	0	0	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$
1	BDI	1	1	1	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$
2	D	0	1	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0	$2\mathbb{Z}$	0
3	DIII	−1	1	1	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0	$2\mathbb{Z}$
4	AII	−1	0	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0	0
5	CII	−1	−1	1	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0	0
6	C	0	−1	0	0	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$	0
7	CI	1	−1	1	0	0	0	$2\mathbb{Z}$	0	$\mathbb{Z}_2$	$\mathbb{Z}_2$	$\mathbb{Z}$

Finally, in Sec. V we will regard **r** as including a temporal variable, and apply the considerations in this paper to classify cyclic pumping processes. The Thouless charge pump[60,61] corresponds to a nontrivial cycle in a system with no symmetries and $\delta=0$ ($d=D=1$). A similar pumping scenario can be applied to superconductors and defines a *fermion parity pump*. This, in turn, is related to the non-Abelian statistics of Ising anyons and provides a framework for understanding braidless operations on systems of three-dimensional superconductors hosting Majorana fermion bound states. Details of several technical calculations can be found in the Appendices. An interesting recent preprint by Freedman *et al.*,[62] which appeared when this manuscript was in its final stages discusses some aspects of the classification of topological defects in connection with a rigorous theory of non-Abelian statistics in higher dimensions.

II. PERIODIC TABLE FOR DEFECT CLASSIFICATION

Table I shows the generalized periodic table for the classification of topological defects in insulators and superconductors. It describes the equivalence classes of Hamiltonians $\mathcal{H}(\mathbf{k},\mathbf{r})$, that can be continuously deformed into one another without closing the energy gap, subject to constraints of particle-hole and/or time-reversal symmetry. These are mappings from a *base space* defined by $(\mathbf{k},\mathbf{r})$ to a *classifying space*, which characterizes the set of gapped Hamiltonians. In order to explain the table, we need to describe (i) the symmetry classes, (ii) the base space, (iii) the classifying space, and (iv) the notion of stable equivalence. The repeating patterns in the table will be discussed in Sec. II C. Much of this section is a review of material in Refs. 35 and 36. What is new is the extension to $D>0$.

A. Symmetry classes

The presence or absence of time reversal symmetry, particle-hole symmetry, and/or chiral symmetry define the ten Altland-Zirnbauer symmetry classes.[44] Time-reversal symmetry implies that

$$\mathcal{H}(\mathbf{k},\mathbf{r}) = \Theta\mathcal{H}(-\mathbf{k},\mathbf{r})\Theta^{-1}, \qquad (2.1)$$

where the antiunitary time reversal operator may be written $\Theta=e^{i\pi S^y/\hbar}K$. S^y is the spin and K is complex conjugation. For spin-1/2 fermions, $\Theta^2=-1$, which leads to Kramers theorem. In the absence of a spin-orbit interaction, the extra invariance of the Hamiltonian under rotations in spin space allows an additional time-reversal operator $\Theta'=K$ to be defined, which satisfies $\Theta'^2=+1$.

Particle-hole symmetry is expressed by

$$\mathcal{H}(\mathbf{k},\mathbf{r}) = -\Xi\mathcal{H}(-\mathbf{k},\mathbf{r})\Xi^{-1}, \qquad (2.2)$$

where Ξ is the antiunitary particle-hole operator. Fundamentally, $\Xi^2=+1$. However, as was the case for Θ, the absence of spin-orbit interactions introduces an additional particle-hole symmetry, which can satisfy $\Xi^2=-1$.

Finally, chiral symmetry is expressed by a unitary operator Π, satisfying

$$\mathcal{H}(\mathbf{k},\mathbf{r}) = -\Pi\mathcal{H}(\mathbf{k},\mathbf{r})\Pi^{-1}. \qquad (2.3)$$

A theory with both particle-hole and time-reversal symmetries automatically has a chiral symmetry $\Pi=e^{i\chi}\Theta\Xi$. The phase χ can be chosen so that $\Pi^2=1$.

Specifying $\Theta^2=0,\pm1$, $\Xi^2=0,\pm1$, and $\Pi^2=0,1$ (here 0 denotes the absence of symmetry) defines the ten Altland-Zirnbauer symmetry classes. They can be divided into two groups: eight *real* classes that have anti unitary symmetries Θ and or Ξ plus two *complex* classes that do not have anti unitary symmetries. Altland and Zirnbauer's notation for these classes, which is based on Cartan's classification of symmetric spaces, is shown in the left-hand part of Table I.

To appreciate the mathematical structure of the eight real symmetry classes it is helpful to picture them on an 8 h "clock," as shown in Fig. 2. The x and y axes of the clock represent the values of Ξ^2 and Θ^2. The "time" on the clock can be represented by an integer s defined modulo 8. Kitaev[36] used a slightly different notation to label the symmetry classes. In his formulation, class D is described by a real Clifford algebra with no constraints, and in the other
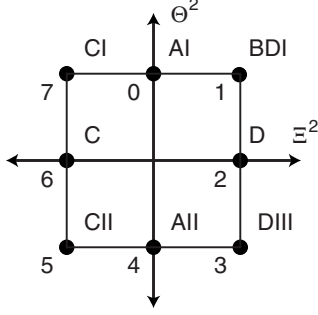
FIG. 2. The eight real symmetry classes that involve the anti-unitary symmetries Θ (time reversal) and/or Ξ (particle hole) are specified by the values of $\Theta^2 = \pm 1$ and $\Xi^2 = \pm 1$. They can be visualized on an eight-hour "clock."

classes Clifford algebra elements are constrained to anticommute with q positive generators. The two formulations are related by $s = q + 2 \mod 8$. The complex symmetry classes can similarly be indexed by an integer s defined modulo 2. For all classes, the presence of chiral symmetry is associated with odd s.

B. Base space, classifying space and stable equivalence

The Hamiltonian is defined on a base space composed of momentum $\mathbf{k}$, defined in a d-dimensional Brillouin zone T^d and real-space degrees of freedom $\mathbf{r}$ in a sphere S^D (or $S^{D-1} \times S^1$ for an adiabatic cycle). The total base space is therefore $T^d \times S^D$ (or $T^d \times S^{D-1} \times S^1$). As in Ref. 36, we will simplify the topological classification by treating the base space as a sphere S^{d+D}. The "strong" topological invariants that characterize the sphere will also characterize $T^d \times S^D$. However, there may be additional topological structure in $T^d \times S^D$ that is absent in S^{d+D}. These correspond to "weak" topological invariants. For $D = 0$ these arise in layered structures. A weak topological insulator, for example, can be understood as a layered two dimensional topological insulator. There are similar layered quantum-Hall states. For $D \neq 0$, then there will also be a weak invariant if the Hamiltonian $\mathcal{H}(\mathbf{k}, \mathbf{r}_0)$ for fixed $\mathbf{r} = \mathbf{r}_0$ is topologically nontrivial. As is the case for the classification of bulk phases $D = 0$, we expect that the topologically protected gapless defect modes are associated with the strong topological invariants.

The set of Hamiltonians that preserve the energy gap separating positive and negative energy states can be simplified without losing any topological information. Consider the retraction of the original Hamiltonian $\mathcal{H}(\mathbf{k}, \mathbf{r})$ to a simpler Hamiltonian whose eigenvalue spectrum is "flattened" so that the positive- and negative-energy states all have the same energy $\pm E_0$. The flattened Hamiltonian is then specified the set of all n eigenvectors [defining a $U(n)$ matrix] modulo unitary rotations within the k conduction bands or the $n - k$ valence bands. The flattened Hamiltonian can thus be identified with a point in the Grassmanian manifold

$$G_{n,k} = U(n)/U(k) \times U(n - k). \tag{2.4}$$

It is useful to broaden the notion of topological equivalence to allow for the presence of extra trivial energy bands.

Two families of Hamiltonians are *stably equivalent* if they can be deformed into one another after adding an arbitrary number of trivial bands. Thus, trivial insulators with different numbers of core energy levels are stably equivalent. Stable equivalence can be implemented by considering an expanded *classifying space* that includes an infinite number of extra conduction and valence bands, $\mathcal{C}_0 = U/U \times U \equiv \cup_{k=0}^{\infty} G_{\infty,k}$.

With this notion of stable equivalence, the equivalence classes of Hamiltonians $\mathcal{H}(\mathbf{k}, \mathbf{r})$ can be formally added and subtracted. The addition of two classes, denoted $[\mathcal{H}_1] + [\mathcal{H}_2]$ is formed by simply combining two independent Hamiltonians into a single Hamiltonian given by the matrix direct sum, $[\mathcal{H}_1 \oplus \mathcal{H}_2]$. Additive inverses are constructed through reversing conduction and valence bands, $[\mathcal{H}_1] - [\mathcal{H}_2] = [\mathcal{H}_1 \oplus -\mathcal{H}_2]$. $[\mathcal{H} \oplus -\mathcal{H}]$ is guaranteed to the be trivial class $[0]$. Because of this property, the stable equivalence classes form an Abelian group, which is the key element of K theory.[63–65]

Symmetries impose constraints on the classifying space. For the symmetry classes with chiral symmetry, Eq. (2.3) restricts $n = 2k$ and the classifying space to a subset $\mathcal{C}_1 = U(\infty) \subset U/U \times U$. The antiunitary symmetries in Eqs. (2.1) and (2.2) impose further constraints. At the special points where $\mathbf{k}$ and $-\mathbf{k}$ coincide, the allowed Hamiltonians are described by the 8 classifying spaces $\mathcal{R}_q$ of real K theory.

C. Properties of the periodic table

For a given symmetry class s, the topological classification of defects is given by the set of stable equivalence classes of maps from the base space $(\mathbf{k}, \mathbf{r}) \in S^{D+d}$ to the classifying space, subject to the symmetry constraints. These form the K group, which we denote as $K_{\mathbb{C}}(s; D, d)$ for the complex symmetry classes and $K_{\mathbb{R}}(s; D, d)$ for the real symmetry classes. These are listed in Table I.

Table I exhibits many remarkable patterns. Many can be understood from the following basic periodicities:

$$K_{\mathbb{F}}(s; D, d + 1) = K_{\mathbb{F}}(s - 1; D, d), \tag{2.5}$$

$$K_{\mathbb{F}}(s; D + 1, d) = K_{\mathbb{F}}(s + 1; D, d). \tag{2.6}$$

Here s is understood to be defined modulo 2 for $\mathbb{F} = \mathbb{C}$ and modulo 8 for $\mathbb{F} = \mathbb{R}$. We will establish these identities mathematically in Appendix A. The basic idea is to start with some Hamiltonian in some symmetry class s and dimensionalities D and d. It is then possible to explicitly construct two new Hamiltonians in one higher dimension which have either (i) $d \to d + 1$ or (ii) $D \to D + 1$. These new Hamiltonians belongs to new symmetry classes that are shifted by 1 "hour" on the symmetry clock and characterized by (i) $s \to s + 1$ or (ii) $s \to s - 1$. We then go on to show that this construction defines a 1–1 correspondence between the equivalence classes of Hamiltonians with the new and old symmetry classes and dimensions, thereby establishing Eqs. (2.5) and (2.6).

The periodicities in Eqs. (2.5) and (2.6) have a number of consequences. The most important for our present purposes is they can be combined to give

$$K_{\mathbb{F}}(s;D+1,d+1) = K_{\mathbb{F}}(s;D,d). \qquad (2.7)$$

This $(1,1)$ periodicity shows that the dependence on the dimensions d and D only occurs via $\delta = d - D$. Thus the dependence of the classifications on D can be deduced from the table for $D=0$. This is one of our central results.

In addition, the periodicities in Eqs. (2.5) and (2.6) explain other features of the table that are already present for $D=0$. In particular, the fact that s is defined modulo 2 (8) for the complex (real) classes leads directly to the Bott periodicity of the dependence of the classifications on d

$$K_{\mathbb{C}}(s;D,d+2) = K_{\mathbb{C}}(s;D,d), \qquad (2.8)$$

$$K_{\mathbb{R}}(s;D,d+8) = K_{\mathbb{R}}(s;D,d). \qquad (2.9)$$

Moreover, Eqs. (2.5) and (2.6) show that $K_a(s;D,d)$ depends only on $d-D-s$. This explains the diagonal pattern in Table I, in which the dependence of the classification on d is repeated in successive symmetry classes. Thus, the entire table could be deduced from a single row.

Equations (2.5) and (2.6) do not explain the pattern of classifications within a single row. Since this is a well-studied math problem there are many routes to the answer.[64,66,67] One approach is to notice that for $d=0$, $K_{\mathbb{F}}(s,D,0)$ is simply the Dth homotopy group of the appropriate classifying space which incorporates the symmetry constraints. For example, for class BDI ($s=1$, $\Xi^2=+1$, and $\Theta^2=+1$) the classifying space is the orthogonal group $O(\infty)$. Then, $K_{\mathbb{R}}(1,D,0) = \pi_D[O(\infty)]$, which are well known. This implies

$$K_{\mathbb{R}}(s;D,d) = \pi_{s+D-d-1}[O(\infty)]. \qquad (2.10)$$

Additional insight can be obtained by examining the interconnections between different elements of the table. For example, the structure within a column can be analyzed by considering the effect of "forgetting" symmetries. Hamiltonians belonging to the real chiral (nonchiral) classes are automatically in complex class AIII (A). There are therefore K group homomorphisms that send any real entries in Table I to complex ones directly above. In particular, as detailed in Appendix B this distinguishes the $\mathbb{Z}$ and $2\mathbb{Z}$ entries, which indicate the possible values of Chern numbers [or $U(n)$ winding numbers] for even (or odd) δ. In addition, the dimensional reduction arguments given in Refs. 10 and 68 lead to a dimensional hierarchy, which helps to explain the pattern within a single row as a function of d.

III. LINE DEFECTS

Line defects can occur at the edge of a 2D system ($\delta = 2 - 0$) or in a 3D system ($\delta = 3 - 1$). From Table I, it can be seen that there are five symmetry classes which can host nontrivial line defects. These are expected to be associated with gapless fermion modes bound to the defect. Table II lists nontrivial classes, along with the character of the associated gapless modes. In the following sections we will discuss each of these cases, along with physical examples.

TABLE II. Symmetry classes that support topologically nontrivial line defects and their associated protected gapless modes.

Symmetry	Topological classes	1D gapless Fermion modes
A	$\mathbb{Z}$	Chiral Dirac
D	$\mathbb{Z}$	Chiral Majorana
DIII	$\mathbb{Z}_2$	Helical Majorana
AII	$\mathbb{Z}_2$	Helical Dirac
C	$2\mathbb{Z}$	Chiral Dirac

A. Class A: Chiral Dirac fermion

1. Topological invariant

A line defect in a generic 3D Bloch band theory with no symmetries is associated with an integer topological invariant. This determines the number of chiral Dirac fermion modes associated with the defect. Since $\mathcal{H}(\mathbf{k},\mathbf{r})$ is defined on a compact four-dimensional space, this invariant is naturally expressed as a second Chern number

$$n = \frac{1}{8\pi^2} \int_{T^3 \times S^1} \text{Tr}[\mathcal{F} \wedge \mathcal{F}], \qquad (3.1)$$

where

$$\mathcal{F} = d\mathcal{A} + \mathcal{A} \wedge \mathcal{A} \qquad (3.2)$$

is the curvature form associated with the non-Abelian Berry's connection $\mathcal{A}_{ij} = \langle u_i | du_j \rangle$ characterizing the valence-band eigenstates $|u_j(\mathbf{k},s)\rangle$ defined on the loop S^1 parameterized by s.

It is instructive to rewrite this as an integral over s of a quantity associated with the local band structure. To this end, it is useful to write $\text{Tr}[\mathcal{F} \wedge \mathcal{F}] = d\mathcal{Q}_3$, where the Chern-Simons 3 form is

$$\mathcal{Q}_3 = \text{Tr}\left[\mathcal{A} \wedge d\mathcal{A} + \frac{2}{3}\mathcal{A} \wedge \mathcal{A} \wedge \mathcal{A}\right]. \qquad (3.3)$$

Now divide the integration volume into thin slices, $T^3 \times \Delta S^1$, where ΔS^1 is the interval between s and $s + \Delta s$. In each slice, Stokes' theorem may be used to write the integral as a surface integral over the surfaces of the slice at s and $s + \Delta s$. In this manner, Eq. (3.1) may be written

$$n = \frac{1}{2\pi} \oint_{S^1} ds \frac{d}{ds} \theta(s), \qquad (3.4)$$

where

$$\theta(s) = \frac{1}{4\pi} \int_{T^3} \mathcal{Q}_3(\mathbf{k},s). \qquad (3.5)$$

Equation (3.5) is precisely the Qi, Hughes, and Zhang formula[10] for the "θ" term that characterizes the magnetoelectric response of a band insulator. $\theta = 0$ for an ordinary time reversal invariant insulator, and $\theta = \pi$ in a strong topological insulator. If parity and time-reversal symmetry are broken then θ can have any intermediate value. We thus conclude that the topological invariant associated with a line
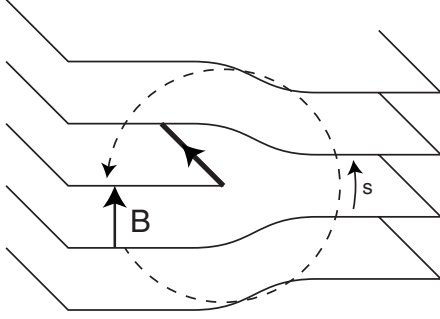
FIG. 3. A line dislocation in a three-dimensional quantum-Hall state characterized by Burgers vector **B**.

defect, which determines the number of chiral fermion branches is given by the winding number of θ. We now consider several examples of 3D line defects that are associated with chiral Dirac fermions.

2. Dislocation in a 3D integer quantum-Hall state

A three-dimensional integer quantum-Hall state can be thought of as a layered version of the two-dimensional integer quantum-Hall state. This can be understood most simply by considering the extreme limit where the layers are completely decoupled 2D systems. A line dislocation, as shown in Fig. 3 will then involve an edge of one of the planes and be associated with a chiral fermion edge state. Clearly, the chiral fermion mode will remain when the layers are coupled, provided the bulk gap remains finite. Here we wish to show how the topological invariant in Eq. (3.1) reflects this fact.

On a loop surrounding the dislocation parameterized by $s \in [0,1]$ we may consider a family of Hamiltonians $H(\mathbf{k},s)$ given by the Hamiltonian of the original bulk crystal displaced by a distance $s\mathbf{B}$, where $\mathbf{B}$ is a lattice vector equal to the Burgers vector of the defect. The corresponding Bloch wave functions will thus be given by

$$u_{m\mathbf{k},s}(\mathbf{r}) = u_{m\mathbf{k}}^0(\mathbf{r} - s\mathbf{B}), \tag{3.6}$$

where $u_{m\mathbf{k}}^0(\mathbf{r})$ are Bloch functions for the original crystal. It then follows that the Berry's connection is

$$\mathcal{A} = \mathcal{A}^0 + \mathbf{B} \cdot [\mathbf{k} - \mathbf{a}^p(\mathbf{k})]ds, \tag{3.7}$$

where

$$\mathcal{A}_{mn}^0(\mathbf{k}) = \langle u_{m\mathbf{k}}^0 | \nabla_{\mathbf{k}} | u_{n\mathbf{k}}^0 \rangle \cdot d\mathbf{k}$$

and

$$\mathbf{a}_{mn}^p(\mathbf{k}) = \langle u_{m\mathbf{k}}^0 | (\nabla_{\mathbf{r}} + \mathbf{k}) | u_{n\mathbf{k}}^0 \rangle. \tag{3.8}$$

With this definition, $\mathbf{a}^p(\mathbf{k})$ is a periodic function: $\mathbf{a}^p(\mathbf{k}+\mathbf{G})$ $=\mathbf{a}^p(\mathbf{k})$ for any reciprocal lattice vector $\mathbf{G}$.[69]

If the crystal is in a three dimensional quantum-Hall state, then the nonzero first Chern number is an obstruction to finding the globally continuous gauge necessary to evaluate Eq. (3.5). We therefore use Eq. (3.1), which can be evaluated by noting that
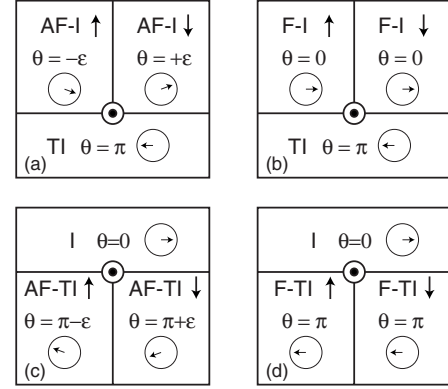
FIG. 4. Heterostructure geometries for chiral Dirac fermions. (a) and (b) show antiferromagnetic or ferromagnetic insulators on the surface of a topological insulator with chiral Dirac fermions at a domain wall. (c) and (d) show a domain wall in an antiferromagnetic or ferromagnetic topological insulator. Chiral fermion modes are present when the domain wall intersects the surface.

$$\mathrm{Tr}[\mathcal{F} \wedge \mathcal{F}] = \mathrm{Tr}[\mathbf{B} \cdot (2\mathcal{F}^0 \wedge d\mathbf{k} - d[\mathcal{F}^0, \mathbf{a}^p]) \wedge ds]. \tag{3.9}$$

Upon integrating $\mathrm{Tr}[\mathcal{F} \wedge \mathcal{F}]$ the total derivative term vanishes due to the periodicity of $\mathbf{a}^p$. Evaluating the integral is then straightforward. The integral over s trivially gives 1. We are then left with

$$n = \frac{1}{2\pi}\mathbf{B} \cdot \mathbf{G}_c, \tag{3.10}$$

where

$$\mathbf{G}_c = \frac{1}{2\pi}\int_{T^3} d\mathbf{k} \wedge \mathrm{Tr}[\mathcal{F}^0]. \tag{3.11}$$

$\mathbf{G}_c$ is a reciprocal lattice vector that corresponds to the triad of Chern numbers that characterize a 3D system. For instance, in a cubic system $\mathbf{G}_c = (2\pi/a)(n_x, n_y, n_z)$, where, for example $n_z = (2\pi)^{-1}\int \mathrm{Tr}[\mathcal{F}_{xy}^0]dk_x \wedge dk_y$, for any value of k_z.

An equivalent formulation is to characterize the displaced crystal in terms of θ. Though Eq. (3.5) cannot be used, Eqs. (3.1) and (3.4) can be used to implicitly define θ up to an arbitrary additive constant

$$\theta(s) = s\mathbf{B} \cdot \mathbf{G}_c. \tag{3.12}$$

3. Topological insulator heterostructures

Another method for engineering chiral Dirac fermions is use heterostructures that combine topological insulators and magnetic materials. The simplest version is a topological insulator coated with a magnetic film that opens a time-reversal symmetry breaking energy gap at the surface. A domain wall is then associated with a chiral fermion mode. In this section we will show how this structure, along with some variants on the theme, fits into our general framework. We first describe the structures qualitatively and then analyze a model that describes them.

Figure 4 shows four possible configurations. Figures 4(a) and 4(b) involve a topological insulator with magnetic mate-

rials on the surface. The magnetic material could be either ferromagnetic or antiferromagnetic. We distinguish these two cases based on whether inversion symmetry is broken or not. Ferromagnetism does not violate inversion symmetry while antiferromagnetism does (at least for inversion about the middle of a bond). This is relevant because θ, discussed above, is quantized unless *both* time reversal and inversion symmetries are violated. Of course, for a noncentrosymmetric crystal inversion is already broken so the distinction is unnecessary.

Figure 4(a) shows a topological insulator capped with antiferromagnetic insulators with $\theta = \pm \epsilon$ separated by a domain wall. Around the junction where the three regions meet θ cycles between π, $+\epsilon$, and $-\epsilon$. Of course this interface structure falls outside the adiabatic regime that Eq. (3.5) is based on. However, it is natural to expect that the physics would not change if the interface was "smoothed out" with θ taking the shortest smooth path connecting its values on either side of the interface.

Figure 4(b) shows a similar device with ferromagnetic insulators, for which $\theta = 0$ or π. In this case the adiabatic assumption again breaks down, however, as emphasized in Ref. 10, the appropriate way to think about the surface is that θ connects 0 and π along a path that is determined by the sign of the induced gap, which in turn is related to the magnetization. In this sense, θ cycles by 2π around the junction.

In Figs. 4(c) and 4(d) we consider topological insulators which have a weak magnetic instability. If in addition to time-reversal, inversion symmetry is broken, then $\theta \sim \pi \pm \epsilon$. Recently Li, Wang, Qi, and Zhang[70] have considered such materials in connection with a theory of a dynamical axion and suggested that certain magnetically doped topological insulators may exhibit this behavior. They referred to such materials as topological magnetic insulators. We prefer to call them *magnetic topological insulators* because as magnetic insulators they are topologically trivial. Rather, they are topological insulators to which magnetism is added. Irrespective of the name, such materials would be extremely interesting to study, and as we discuss below, may have important technological utility.

Figure 4(c) shows two antiferromagnetic topological insulators with $\theta = \pi \pm \epsilon$ separated by a domain wall, and Fig. 4(d) shows a similar device with ferromagnetic topological insulators. They form an interface with an insulator, which could be vacuum. Under the same continuity assumptions as above the junction where the domain wall meets the surface will be associated with a chiral fermion mode. Like the structure in Fig. 4(a), this may be interpreted as an edge state on a domain wall between the "half-quantized" quantum-Hall states of the topological insulator surfaces. However, an equally valid interpretation is that the domain wall itself forms a single two-dimensional integer quantum-Hall state with an edge state. Our framework for topologically classifying the line defects underlies the equivalence between these two points of view.

Mong, Essen, and Moore[71] have introduced a *different* kind of antiferromagnetic topological insulator that relies on the symmetry of time reversal combined with a lattice translation. Due to the necessity of translation symmetry, however, such a phase is not robust to disorder. They found that

chiral Dirac modes occur at certain step edges in such crystals. These chiral modes can also be understood in terms of the invariant in Eq. (3.1). Note that these chiral modes survive in the presence of disorder even though the bulk state does not. Thus, the chiral mode, protected by the strong invariant in Eq. (3.1), is more robust than the bulk state that gave rise to it.

If one imagines weakening the coupling between the two antiferromagnetic topological insulators (using our terminology, not that of Mong, *et al.*[71]) and taking them apart, then at some point the chiral mode has to disappear. At that point, rather than taking the "shortest path" between $\pi \pm \epsilon$, θ takes a path that passes through 0. At the transition between the "short-path" and the "long-path" regimes, the gap on the domain wall must go to zero, allowing the chiral mode to escape. This will have the character of a plateau transition in the 2D integer quantum-Hall effect.

Structures involving magnetic topological insulators would be extremely interesting to study because with them it is possible to create chiral fermion states with a single material. Indeed, one can imagine scenarios where a magnetic memory, encoded in magnetic domains, could be read by measuring the electrical transport in the domain wall chiral fermions.

To model the chiral fermions in these structures we begin with the simple three-dimensional model for trivial and topological insulators considered in Ref. 10

$$\mathcal{H}_0 = v\mu_x \vec{\sigma} \cdot \mathbf{k} + (m + \epsilon |\mathbf{k}|^2)\mu_z. \tag{3.13}$$

Here $\vec{\sigma}$ represents spin and μ_z describes an orbital degree of freedom. $m > 0$ describes the trivial insulator and $m < 0$ describes the topological insulator. An interface where m changes sign is then associated with gapless surface states.

Next consider time-reversal symmetry-breaking perturbations, which could arise from exchange fields due to the presence of magnetic order. Two possibilities include

$$\mathcal{H}_{af} = h_{af}\mu_y, \tag{3.14}$$

$$\mathcal{H}_f = \vec{h}_f \cdot \vec{\sigma}. \tag{3.15}$$

Either h_{af} or $h_{f,z}$ will introduce a gap in the surface states but they have different physical content. $\mathcal{H}_0$ has an inversion symmetry given by $\mathcal{H}_0(\mathbf{k}) = P\mathcal{H}_0(-\mathbf{k})P$ with $P = \mu_z$. Clearly, $\mathcal{H}_f$ respects this inversion symmetry. $\mathcal{H}_{af}$ does not respect P but does respect $P\Theta$. We therefore associate $\mathcal{H}_f$ with ferromagnetic order and $\mathcal{H}_{af}$ with antiferromagnetic order.

Within the adiabatic approximation, the topological invariant in Eq. (3.1) can be evaluated in the presence of either Eq. (3.14) or (3.15). The antiferromagnetic perturbation in Eq. (3.14) is most straightforward to analyze because $\mathcal{H}_0 + H_{af}$ is a combination of five anticommuting Dirac matrices. On a circle surrounding the junction parameterized by s it can be written in the general form

$$H(\mathbf{k}, s) = \mathbf{h}(\mathbf{k}, s) \cdot \vec{\gamma}, \tag{3.16}$$

where $\vec{\gamma} = (\mu_x \sigma_x, \mu_x \sigma_y, \mu_x \sigma_z, \mu_z, \mu_y)$ and $\mathbf{h}(\mathbf{k}, s) = [v\mathbf{k}, m(s) + \epsilon |\mathbf{k}|^2, h_{af}(s)]$. For a model of this form, the second Chern number in Eq. (3.1) is given simply by the winding number

of the unit vector $\hat{\mathbf{d}}(\mathbf{k},s)=\mathbf{h}/|\mathbf{h}| \in S^4$ as a function of $\mathbf{k}$ and s. This is most straightforward to evaluate in the limit $\epsilon \to 0$, where $\hat{\mathbf{d}}$ is confined to the "equator" $(d_1,d_2,d_3,0,0)$ everywhere except near $\mathbf{k} \sim 0$ and $|\mathbf{k}| \gtrsim 1/\epsilon$. The winding number is determined by the behavior at $\mathbf{k} \sim 0$, and may be expressed by Eq. (3.4) with θ given by

$$e^{i\theta} = \frac{m + ih_{af}}{\sqrt{m^2 + h_{af}^2}}. \tag{3.17}$$

We therefore expect a topological line defect to occur at an intersection between planes where m and h_{af} change sign. The chiral fermion mode associated with this defect can seen explicitly if we solve a simple linear model, $m=f_z z$, $h_{af}=f_y y$. This model, which has the form of a harmonic oscillator, is solved in Appendix C, and explicitly gives the chiral Dirac fermion mode with dispersion

$$E(k_x) = v \, \mathrm{sgn}(f_z f_y)k_x. \tag{3.18}$$

B. Class D: Chiral Majorana fermions

1. Topological invariant

A line defect in a superconductor without time-reversal symmetry is characterized by an integer topological invariant that determines the number of associated chiral Majorana fermion modes. Since the BdG Hamiltonian characterizing a superconductor has the same structure as the Bloch Hamiltonian, we can analyze the problem by "forgetting" about the particle-hole symmetry and treating the BdG Hamiltonian as if it was a Bloch Hamiltonian. The second Chern number, given by Eq. (3.1) can be defined. It can be verified that any value of the Chern number is even under particle-hole symmetry so that particle-hole symmetry does not rule out a nonzero Chern number. We may follow the same steps as Eqs. (3.1)–(3.5) to express the integer topological invariant as

$$\tilde{n} = \frac{1}{8\pi^2} \int_{T^3 \times S^1} \mathrm{Tr}[\tilde{\mathcal{F}} \wedge \tilde{\mathcal{F}}], \tag{3.19}$$

where $\tilde{\mathcal{F}}$ is the curvature form characterizing the BdG theory. As in Eq. (3.4), $\tilde{n}$ may be expressed as a winding number of $\tilde{\theta}$, which is expressed as an integral over the Brillouin zone of the Chern-Simons 3 form. The difference between n and $\tilde{n}$ is that $\tilde{n}$ characterizes a BdG Hamiltonian. If we considered the BdG Hamiltonian for a nonsuperconducting insulator, then due to the doubling in the BdG equation, we would find

$$\tilde{n} = 2n. \tag{3.20}$$

In this case, the chiral Dirac fermion that occurs for a 2π ($n=1$) winding of θ corresponds to a 4π ($n=2$) winding of $\tilde{\theta}$. Superconductivity allows for the possibility of a 2π winding in $\tilde{\theta}$: a chiral Dirac fermion can be split into a pair of chiral Majorana fermions.
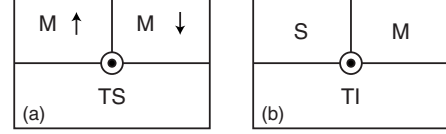
FIG. 5. Heterostructure geometries for Chiral Majorana fermions. (a) shows a magnetic domain wall on the surface of a topological superconductor while (b) shows an interface between a superconductor and a magnet on the surface of a topological insulator.

2. Dislocation in a layered topological superconductor

The simplest example to consider is a dislocation in a three-dimensional superconductor. The discussion closely parallels Sec. III A 2 and we find

$$\tilde{n} = \frac{1}{2\pi}\mathbf{B} \cdot \tilde{\mathbf{G}}_c, \tag{3.21}$$

where $\mathbf{B}$ is the Burgers vector of the dislocation and $\tilde{\mathbf{G}}_c$ characterizes the triad of first Chern numbers characterizing the 3D BdG Hamiltonian. A 3D system consisting of layers of a 2D topological superconductor will be characterized by a nonzero $\tilde{\mathbf{G}}_c$. Since, as a 3D superconductor, the layered structure is in the topologically trivial class, such a state could be referred to as a *weak topological superconductor*.

The simplest model system in this class is a stack of 2D p_x+ip_y superconductors. A dislocation would then have $\tilde{n}=1$ and a single chiral Majorana fermion branch. A possible physical realization of the weak topological superconductor state is Sr_2RuO_4, which may exhibit triplet p_x+ip_y pairing. Since the spin-up and spin-down electrons make two copies of the spinless state, a dislocation will be associated with $\tilde{n}=2$. Thus, we predict that there will be two chiral Majorana modes bound to the dislocation, which is the same as a single chiral Dirac fermion mode.

3. Superconductor heterostructures

We now consider heterostructures with associated chiral Majorana modes. The simplest to consider is a BdG analog of the structures considered in Fig. 5. These would involve, for example, an interface between a 3D time-reversal invariant topological superconductor with a magnetic material with a magnetic domain wall. The analysis of such a structure is similar to that in Eq. (3.13) if we replace the Pauli matrices describing the orbital degree of freedom $\vec{\mu}$ with Pauli matrices describing Nambu space $\vec{\tau}$. Protected chiral Majorana fermion modes of this sort on the surface of ^{3}He-B with a magnetic domain wall have been recently discussed by Volovik.[72]

In Ref. 48 a different method for engineering chiral Majorana fermions was introduced by combining an interface between superconducting and magnetic regions on the surface of a topological insulator. To describe this requires the eight-band model introduced in Ref. 53

$$H = \tau_z \mu_x \vec{\sigma} \cdot \mathbf{k} + (m + \epsilon|\mathbf{k}|^2)\tau_z\mu_z + \Delta\tau_x + h\mu_y. \tag{3.22}$$

(Here, for simplicity we consider only the antiferromagnetic term). The surface of the topological insulator occurs at a

domain wall (say, in the x-y plane), where $m(z)$ changes sign. The superconducting order parameter Δ and magnetic perturbation h both lead to an energy gap in the surface states. This Hamiltonian is straightforward to analyze because $[\mathcal{H}, \tau_x \mu_y] = 0$, which allows the 8×8 problem to be divided into two 4×4 problems, which have superconducting/magnetic mass terms $\Delta \pm h$. Near a defect where $\Delta = h$ the $\Delta + h$ gap never closes while the $\Delta - h$ gap can be critical. $\Delta > h$ leads to a superconducting state while $\Delta < h$ leads to a quantum-Hall-type state. There is a transition between the two at $\Delta = h$.

An explicit model for the line defect can be formulated with $m(z) = f_z z$, $\Delta - h = f_y y$ and $\Delta + h = M$. The topological invariant in Eq. (3.19) can be evaluated using a method similar to Eq. (3.16) and the chiral Majorana states can be explicitly solved along the lines of Eq. (3.18).

C. Class AII: Helical Dirac fermions

1. Topological invariant

Line defects for class AII are characterized by a $\mathbb{Z}_2$ topological invariant. To develop a formula for this invariant we follow the approach used in Ref. 73 to describe the invariant characterizing the quantum spin-Hall insulator.

As in the previous section, a line defect in three dimensions is associated with a four parameter space $(\mathbf{k}, \mathbf{r}) \in T^3 \times S^1$. Due to time-reversal symmetry, the second Chern number that characterized the line defects in Eq. (3.1) must be zero. Thus there is no obstruction to defining Bloch basis functions $|u(\mathbf{k}, \mathbf{r})\rangle$ continuously over the entire base space. However, the time reversal relation between $(-\mathbf{k}, \mathbf{r})$ and $(\mathbf{k}, \mathbf{r})$ allows for an additional constraint so that the state is specified by the degrees of freedom in *half* the Brillouin zone.

As in Ref. 73 it is useful to define a matrix

$$w_{mn}(\mathbf{k}, \mathbf{r}) = \langle u_m(\mathbf{k}, \mathbf{r}) | \Theta | u_n(-\mathbf{k}, \mathbf{r}) \rangle. \qquad (3.23)$$

Because $|u_m(\mathbf{k}, \mathbf{r})\rangle$ and $|u_n(-\mathbf{k}, \mathbf{r})\rangle$ are related by time-reversal symmetry $w(\mathbf{k}, \mathbf{r})$ is a unitary matrix that depends on the gauge choice for the basis functions. Locally it is possible to choose a basis in which

$$w(\mathbf{k}, \mathbf{r}) = w_0, \qquad (3.24)$$

where w_0 is independent of $\mathbf{k}$ and $\mathbf{r}$ so that states at $(\pm \mathbf{k}, \mathbf{r})$ have a fixed relation. Since for $\mathbf{k} = 0$ $w = -w^T$, w_0 must be antisymmetric. A natural choice is thus $w_0 = i\sigma_2 \otimes 1$.

The $\mathbb{Z}_2$ topological invariant is an obstruction to finding such a constrained basis globally. The constrained basis can be defined on two patches but the basis functions on the two patches are necessarily related by a topologically nontrivial transition function. In this sense, the $\mathbb{Z}_2$ invariant resembles the second Chern number in Eq. (3.1).

In Appendix E we will generalize the argument developed in Ref. 73 to show that the transition function relating the two patches defines the $\mathbb{Z}_2$ topological invariant, which may be written[74]

$$\nu = \frac{1}{8\pi^2} \left(\int_{(1/2)T^3 \times S^1} \mathrm{Tr}[\mathcal{F} \wedge \mathcal{F}] - \int_{\partial (1/2)T^3 \times S^1} \mathcal{Q}_3 \right) \mod 2,$$

$$(3.25)$$

where $\mathcal{F}$ and $\mathcal{Q}_3$ are expressed in terms of the Berry's connection $\mathcal{A}$ using Eqs. (3.2) and (3.3). The integral is over half of the base space $(1/2)(T^3 \times S^1)$, defined such that $(\mathbf{k}, \mathbf{r})$ and $(-\mathbf{k}, \mathbf{r})$ are never both included. The second term is over the boundary of $(1/2)(T^3 \times S^1)$, which is closed under $(\mathbf{k}, \mathbf{r}) \rightarrow (-\mathbf{k}, \mathbf{r})$. Equation (3.25) must be used with care because the Chern Simons form in the second term depends on the gauge. A different continuous gauge can give a different ν, but due to Eq. (3.24), they must be related by an even integer. Thus, an odd number is distinct.

In addition to satisfying Eq. (3.24), it is essential to use a gauge in which at least $\mathcal{Q}_3$ is continuous on $\partial \frac{1}{2} T^3 \times S^1$ (though not necessarily on all of $\frac{1}{2} T^3 \times S^1$). This continuous gauge can always be found if the base space is a sphere S^4. However for $T^3 \times S^1$, the weak topological invariants can pose an obstruction to finding a continuous gauge. We will show how to work around this difficulty at the end of the following section.

2. Dislocation in a weak topological insulator

Ran, Zhang, and Vishwanath recently studied the problem of a line dislocation in a topological insulator.[75] They found that an insulator with nontrivial weak topological invariants can exhibit topologically protected helical modes at an appropriate line dislocation. In this section we will show that these protected modes are associated with a nontrivial $\mathbb{Z}_2$ invariant in Eq. (3.25). In addition to providing an explicit example for this invariant, this formulation provides additional insight into why protected modes can exist in a weak topological insulator. As argued in Ref. 9 and 12, the weak topological invariants lose their meaning in the presence of disorder. The present considerations show that the helical modes associated with the dislocation are protected by the *strong* topological invariant associated with the line defect. Thus if we start with a perfect crystal and add disorder, then the helical modes remain, even though the crystal is no longer a weak topological insulator. The helical modes remain even if the disorder destroys the crystalline order, so that dislocations become ill defined, *provided* the mobility gap remains finite in the bulk crystal. In this case, the Hamiltonian has a nontrivial winding around the line defect, even though the defect has no obvious structural origin. Thus, the weak topological insulator provides a *route* to realizing the topologically protected line defect. But once present, the line defect is more robust than the weak topological insulator.

To evaluate the $\mathbb{Z}_2$ invariant in Eq. (3.25) for a line dislocation we repeat the analysis in Sec. III A 2. Because of the subtlety with the application of Eq. (3.25) we will first consider the simplest case of a dislocation in a weak topological insulator. Afterward we will discuss the case of a crystal with both weak and strong invariants.

The Bloch functions on a circle surrounding a dislocation are described by Eq. (3.6) and the evaluation of $\mathrm{Tr}[\mathcal{F} \wedge \mathcal{F}]$ proceeds exactly as in Eqs. (3.7)–(3.9). To evaluate the sec-

ond term in Eq. (3.25) we need the Chern-Simons 3 form. One approach is to use Eqs. (3.3) and (3.7). However, this is not continuously defined on $\partial(1/2)(T^3 \times S^1)$ because $\mathcal{A}$ has a term $\mathbf{B} \cdot \mathbf{k} ds$ that is discontinuous at the Brillouin zone boundary. An alternative is to write

$$\mathcal{Q}_3 = \mathrm{Tr}[\mathbf{B} \cdot (2\mathcal{A}^0 \wedge d\mathbf{k} - [\mathcal{F}^0, \mathbf{a}^p]) \wedge ds]. \qquad (3.26)$$

From Eq. (3.9) this clearly satisfies $\mathrm{Tr}[\mathcal{F} \wedge \mathcal{F}] = d\mathcal{Q}_3$ and it is defined continuously on $\partial(1/2)(T^3 \times S^1)$ as long as $\mathcal{A}^0$ is continuously defined on $\partial(1/2)T^3$. For a weak topological insulator this is always possible, provided $(1/2)T^3$ is defined appropriately. Equation (3.26) differs from Eq. (3.3) by a total derivative.

Combining Eqs. (3.9), (3.25), and (3.26), the terms involving $\mathbf{a}^p$ cancel because $\mathbf{a}^p$ is globally defined. (Note that $\mathbf{a}^p$ is unchanged by a $\mathbf{k}$-dependent—but $\mathbf{r}$-independent—gauge transformation). This cannot be said of the term involving $\mathcal{A}^0$, however, because in a weak topological insulator $\mathcal{A}^0$ is *not* globally defined on $(1/2)T^3$. Performing the trivial integral over s we then find

$$\nu = \frac{1}{2\pi} \mathbf{B} \cdot \mathbf{G}_\nu \bmod 2, \qquad (3.27)$$

where

$$\mathbf{G}_\nu = \int_{1/2 T^3} \mathrm{Tr}[\mathcal{F}^0] \wedge d\mathbf{k} - \int_{\partial 1/2 T^3} \mathrm{Tr}[\mathcal{A}^0] \wedge d\mathbf{k}.$$

$$(3.28)$$

The simplest case to consider is a weak topological insulator consisting of decoupled layers of 2D quantum spin-Hall insulator stacked with a lattice constant a in the z direction. In this case $\mathcal{F}^0 = \mathcal{F}^0(k_x, k_y)$ is independent of k_z so the k_z integral can be performed trivially. This leads to $\mathbf{G}_\nu = (2\pi/a)\nu\hat{\mathbf{z}}$, where

$$\nu = \frac{i}{2\pi} \left[\int_{1/2 T^2} \mathrm{Tr}[\mathcal{F}^0] - \int_{\partial 1/2 T^2} \mathrm{Tr}[\mathcal{A}^0] \right] \qquad (3.29)$$

is the 2D $\mathbb{Z}_2$ topological invariant characterizing the individual layers.

Equation (3.28) also applies to a more general 3D weak topological insulator. A weak topological insulator is characterized by a triad of $\mathbb{Z}_2$ invariants $(\nu_1\nu_2\nu_3)$ that define a mod 2 reciprocal lattice vector[9,12]

$$\mathbf{G}_\nu = \nu_1 \mathbf{b}_1 + \nu_2 \mathbf{b}_2 + \nu_3 \mathbf{b}_3, \qquad (3.30)$$

where $\mathbf{b}_i$ are primitive reciprocal lattice vectors corresponding to primitive lattice vectors $\mathbf{a}_i$ (such that $\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}$). The indices ν_i can be determined by evaluating the 2D invariant in Eq. (3.29) on the time-reversal invariant plane $\mathbf{k} \cdot \mathbf{a}_i = \pi$.

To show that $\mathbf{G}_\nu$ in Eqs. (3.28) and (3.30) are equivalent, consider $\mathbf{G}_\nu \cdot \mathbf{a}_1$ in Eq. (3.28). If we write $\mathbf{k} = x_1 \mathbf{b}_1 + x_2 \mathbf{b}_2 + x_3 \mathbf{b}_3$, then the integrals over x_2 and x_3 have the form of Eq. (3.29). Since this is quantized, it must be independent of x_1 and will be given by its value at $x_1 = 1/2$. This then gives $\mathbf{G}_\nu \cdot \mathbf{a}_1 = 2\pi\nu_1$. A similar analysis of the other components establishes the equivalence. A nontrivial value of Eq. (3.27)

is the same as the criterion for the existence of protected helical modes on a dislocation Ran, Zhang, and Vishwanath[75] derived using a different method.

Evaluating Eq. (3.28) in a crystal that is *both* a strong topological insulator and a weak topological insulator (such as $\mathrm{Bi}_{1-x}\mathrm{Sb}_x$) is problematic because the 2D invariants evaluated on the planes $x_1 = 0$ and $x_1 = 1/2$ are necessarily *different* in a strong topological insulator. This arises because a nontrivial strong topological invariant ν_0 is an obstruction to continuously defining $\mathcal{A}^0$ on $\partial(1/2)T^3$ so Eq. (3.29) cannot be evaluated continuously between $x_1 = 0$ and $x_1 = 1/2$. From the point of view of the topological classification of the *defect* on $T^3 \times S^1$, ν_0 is like a *weak* topological invariant because it a property of T^3 and is independent of the real-space parameter s in S^1. Thus this complication is a manifestation of the fact that topological classification of Hamiltonians on $T^3 \times S^1$ has more structure than those on S^4. The problem is not with the existence of the invariant ν on $T^3 \times S^1$ but rather with applying the formulas (3.25) and (3.28). The problem can be circumvented with the following trick.

Consider an auxiliary Hamiltonian $\widetilde{\mathcal{H}}(\mathbf{k}, \mathbf{r}) = \mathcal{H}(\mathbf{k}, \mathbf{r}) \oplus \mathcal{H}_{STI}(\mathbf{k})$, where $\mathcal{H}_{STI}$ is a simple model Hamiltonian for a strong topological insulator such as Eq. (3.13), which can be chosen such that it is a constant independent of $\mathbf{k}$ everywhere except in a small neighborhood close to $\mathbf{k} = 0$ where a band inversion occurs. Adding such a Hamiltonian that is independent of $\mathbf{r}$ will have no effect on the topologically protected modes associated with a line defect so we expect the invariant ν to be the same for both $\mathcal{H}(\mathbf{k}, \mathbf{r})$ and $\widetilde{\mathcal{H}}(\mathbf{k}, \mathbf{r})$. If $\mathcal{H}(\mathbf{k}, \mathbf{r})$ has a nontrivial strong topological invariant $\nu_0 = 1$ then $\widetilde{\mathcal{H}}(\mathbf{k}, \mathbf{r})$ will have $\nu_0 = 0$, so that Eq. (3.28) can be applied. $\mathbf{G}_\nu$ will then be given by the 2D invariant in Eq. (3.29) evaluated for $\widetilde{\mathcal{H}}$, which will be independent of x_1. Since $\mathcal{H}_{STI}(\mathbf{k})$ is $\mathbf{k}$-independent everywhere except a neighborhood of $\mathbf{k} = 0$, this will agree with the 2D invariant evaluated for $\mathcal{H}$ at $x_1 = 1/2$, but not $x_1 = 0$. It then follows that even in a strong topological insulator the invariant characterizing a line dislocation is given by Eq. (3.27), where $\mathbf{G}_\nu$ is given by Eq. (3.30) in terms of the weak topological invariants.

3. Heterostructure geometries

In principle, it may be possible to realize 1D helical fermions in a 3D system that does not rely on a weak topological insulating state. It is possible to write down a 3D model, analogous to Eq. (3.13) that has bound helical modes. However, it is not clear how to physically implement this model. This model will appear in a more physical context as a BdG theory in the following section.

D. Class DIII: Helical Majorana fermions

Line defects for class DIII are characterized by a $\mathbb{Z}_2$ topological invariant that signals the presence or absence of 1D helical Majorana fermion modes. As in Sec. III B, the BdG Hamiltonian has the same structure as a Bloch Hamiltonian, and the $\mathbb{Z}_2$ invariant can be deduced by "forgetting" the particle-hole symmetry, and treating the problem as if it was a Bloch Hamiltonian in class AII.
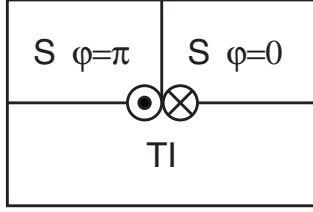
FIG. 6. Helical Majorana fermions at a linear Josephson junction with phase difference π on the surface of a topological insulator.

There are several ways to realize helical Majorana fermions. The simplest is to consider the edge of a 2D time-reversal invariant superconductor or superfluid, or equivalently a dislocation in a layered version of that 2D state. A second is to consider a topological line defect in a 3D class DIII topological superconductor or superfluid. Such line defects are well known in of ^{3}He B (Refs. 41 and 76) and have recently been revisited in Refs. 37 and 77.

Here we will consider a different realization that uses topological insulators and superconductors. Consider a linear junction between two superconductors on the surface of a topological insulator as shown in Fig. 6. In Ref. 48 it was shown that when the phase difference between the superconductors is π there are gapless helical Majorana modes that propagate along the junction. This can be described by an eight-band minimal model that describes a topological insulator surface with a superconducting proximity effect

$$\mathcal{H} = v\tau_z\mu_x\vec{\sigma}\cdot\mathbf{k} + (m+\epsilon|\mathbf{k}|^2)\tau_z\mu_z + \Delta_1\tau_x. \quad (3.31)$$

Here m is the mass describing the band inversion of a topological insulator, as in Eq. (3.22), and Δ_1 is the real part of the superconducting gap parameter. This model has time reversal symmetry with $\Theta = i\sigma_y K$ and particle-hole symmetry with $\Xi = \sigma_y\tau_y K$. The imaginary part of the superconducting gap, $\Delta_2\tau_y$ violates time-reversal symmetry. A line junction along the x direction with phase difference π at the surface of a topological insulator corresponds to the intersection of planes where $m(z)$ and $\Delta_1(y)$ change sign.

The $\mathbb{Z}_2$ invariant characterizing such a line defect is straightforward to evaluate because $[\mathcal{H},\mu_y\tau_x]=0$. This extra symmetry allows a "spin Chern number" to be defined, $n_\sigma = (16\pi^2)^{-1}\int\mathrm{Tr}[\mu_y\tau_x\mathcal{F}\wedge\mathcal{F}]$. Since the system decouples into two time reversed versions of Eq. (3.13), $n_\sigma=1$. By repeating the formulation in Appendix E of the $\mathbb{Z}_2$ invariant ν, it is straightforward to show that this means $\nu=1$.

The helical modes can be explicitly seen by solving the linear theory, $m=f_z z$, $\Delta_1=f_y y$, which leads to the harmonic oscillator model studied in Appendix C. In the space of the two zero modes the Hamiltonian has the form

$$\mathcal{H} = vk_x\sigma_x \quad (3.32)$$

and describes 1D helical Majorana fermions.

E. Class C: Chiral Dirac fermions

We finally briefly consider line defects in class C. Class C can be realized when time-reversal symmetry is broken in a

superconductor without spin orbit interactions that has even parity singlet pairing. Line defects are characterized by an integer topological invariant that determines the number of chiral Majorana fermion modes associated with the line. As in class D, this may be evaluated by forgetting the particle-hole symmetry and evaluating the corresponding Chern number that would characterize class A. The $2\mathbb{Z}$ in Table II for this case, however, means that the Chern integer computed in this manner is necessarily *even*. This means that there will necessarily be an even number $2n$ of chiral Majorana fermion modes, which may equivalently viewed as n chiral Dirac fermion modes. An example of such a system would be a 2D $d_{x^2-y^2}+id_{xy}$ superconductor,[78] which exhibits chiral Dirac fermion edge states, or equivalently a dislocation in a 3D layered version of that state.

IV. POINT DEFECTS

Point defects can occur at the end of a 1D system ($\delta=1$ -0) or at topological defects in 2D ($\delta=2-1$) or 3D ($\delta=3$ -2) systems. From the $\delta=1$ column Table I, it can be seen that there are five symmetry classes that can have topologically nontrivial point defects. These are expected to be associated with protected zero-energy bound states. Table III lists the nontrivial classes, along with the character of the associated zero modes. In this section we will discuss each of these cases.

A. Classes AIII, BDI, and CII: Chiral zero modes

1. Topological invariant and zero modes

Point defects in classes AIII, BDI, and CII are characterized by integer topological invariants. The formula for this integer invariant can be formulated by exploiting the chiral symmetry in each class. In a basis where the chiral symmetry operator is $\Pi=\tau_z$, the Hamiltonian may be written

$$\mathcal{H}(\mathbf{k},\mathbf{r}) = \begin{pmatrix} 0 & q(\mathbf{k},\mathbf{r}) \\ q(\mathbf{k},\mathbf{r})^\dagger & 0 \end{pmatrix}. \quad (4.1)$$

When the Hamiltonian has a flattened eigenvalue spectrum $\mathcal{H}^2=1$, $q(\mathbf{k},\mathbf{r})$ is a unitary matrix. For a point defect in d dimensions, the Hamiltonian as a function of d momentum variables and $D=d-1$ position variables is characterized by

TABLE III. Symmetry classes supporting nontrivial point topological defects and their associated $E=0$ modes.

Symmetry	Topological classes	$E=0$ bound states
AIII	$\mathbb{Z}$	Chiral Dirac
BDI	$\mathbb{Z}$	Chiral Majorana
D	$\mathbb{Z}_2$	Majorana
DIII	$\mathbb{Z}_2$	Majorana Kramers doublet (=Dirac)
CII	$2\mathbb{Z}$	Chiral Majorana Kramers Doublet (=Chiral Dirac)

the winding number associated with the homotopy $\pi_{2d-1}[U(n \to \infty)] = \mathbb{Z}$, which is given by

$$n = \frac{(d-1)!}{(2d-1)!(2\pi i)^d} \int_{T^d \times S^{d-1}} \text{Tr}[(q dq^\dagger)^{2d-1}]. \quad (4.2)$$

For a Hamiltonian that is built from anticommuting Dirac matrices, $\mathcal{H}(\mathbf{k},\mathbf{r}) = \hat{\mathbf{d}}(\mathbf{k},\mathbf{r}) \cdot \tilde{\gamma}$, this invariant is given simply by the winding degree of the mapping $\hat{\mathbf{d}}(\mathbf{k},\mathbf{r})$ from $T^d \times S^{d-1}$ to S^{2d-1}, which is expressed as an integral of the Jacobian

$$n = \frac{(d-1)!}{2\pi^d} \int_{T^d \times S^{d-1}} d^d \mathbf{k} d^{d-1} \mathbf{r} \frac{\partial \hat{\mathbf{d}}(\mathbf{k},\mathbf{r})}{\partial^d \mathbf{k} \partial^{d-1} \mathbf{r}}. \quad (4.3)$$

In class AIII there are no constraints on $q(\mathbf{k},\mathbf{r})$ other than unitarity so all possible values of n are possible. There are additional constraints for the chiral classes with antiunitary symmetries. As shown in Appendix B, this is simplest to see by analyzing the constraints on the winding degree discussed above. n must be zero in classes CI and DIII. There is no constraint on n in class BDI while n must be even in class CII.

The topological invariant is related to an index that characterizes the chirality of the zero modes

$$n = N_+ - N_-, \quad (4.4)$$

where $N_\pm$ are the number of zero modes that are eigenstates of Π with eigenvalue ± 1. To see that these zero modes are indeed protected consider $N_+ = n > 0$ and $N_- = 0$. Any term in the Hamiltonian that could shift any of the N_+ degenerate states would have to have a nonzero matrix element connecting states with the same chirality. Such terms are forbidden, though, by the chiral symmetry $\{\mathcal{H}, \Pi\} = 0$. In the superconducting classes BDI and CII the zero energy states are Majorana bound states. In class CII, however, since time reversal symmetry requires that n must be even, the paired Majorana states can be regarded as zero-energy Dirac fermion states.

In the special case where $\mathcal{H}(\mathbf{k},\mathbf{r})$ has the form of a massive Dirac Hamiltonian, by introducing a suitable regularization for $|\mathbf{k}| \to \infty$ the topological invariant in Eqs. (4.2) and (4.3) can be expressed in a simpler manner as a topological invariant characterizing the mass term. In the following sections we consider this in the three specific cases $d = 1, 2, 3$.

2. Solitons in d = 1

The simplest topological zero mode occurs in the Jackiw-Rebbi model,[45] which is closely related to the Su, Schrieffer, and Heeger model.[47] Consider

$$\mathcal{H}(k,x) = v k \sigma_x + m \sigma_y. \quad (4.5)$$

Domain walls where $m(x)$ changes sign as a function of x are associated with the well known zero-energy soliton states.

To analyze the topological class requires a regularization for $|k| \to \infty$. This can either be done with a lattice, as in the Su, Schrieffer, Heeger model or by adding a term $\epsilon k^2 \sigma_y$, as in Eq. (3.13) so that $|k| \to \infty$ can be replaced by a single point.

In either case, the invariant in Eq. (4.2) changes by 1 when m changes sign.

3. Jackiw-Rossi Model in d = 2

Jackiw and Rossi introduced a two-dimensional model that has protected zero modes.[46] The Hamiltonian can be written

$$\mathcal{H}(\mathbf{k},\mathbf{r}) = v \tilde{\gamma} \cdot \mathbf{k} + \vec{\Gamma} \cdot \vec{\phi}(\mathbf{r}), \quad (4.6)$$

where $\mathbf{k} = (k_x, k_y)$, and (γ_1, γ_2) and (Γ_1, Γ_2) are anticommuting Dirac matrices. They showed that the core of a vortex where $\phi = \phi_1 + i\phi_2$ winds by $2\pi n$ is associated with n zero modes that are protected by the chiral symmetry. Viewed as a BdG Hamiltonian, these zero modes are Majorana bound states.

This can be interpreted as a Hamiltonian describing superconductivity in Dirac fermions. In this interpretation the Dirac matrices are expressed as $(\gamma_1, \gamma_2) = \tau_z(\sigma_x, \sigma_y)$ and $(\Gamma_1, \Gamma_2) = (\tau_x, \tau_y)$, where $\vec{\sigma}$ is a Pauli matrix describing spin and $\vec{\tau}$ describes particle-hole space. The superconducting pairing term is $\Delta = \phi_1 + i\phi_2$. In this interpretation a vortex violates the physical time-reversal symmetry $\Theta = i\sigma_y K$. However, even in the presence of a vortex this model has a fictitious "time-reversal symmetry" $\tilde{\Theta} = \sigma_x \tau_x K$ which satisfies $\tilde{\Theta}^2 = +1$. This symmetry would be violated by a finite chemical potential term $\mu \tau_z$. Combined with particle-hole symmetry $\tilde{\Xi} = \sigma_y \tau_y K$ ($\tilde{\Xi}^2 = +1$), $\tilde{\Theta}$ defines the BDI class with chiral symmetry $\Pi = \sigma_z \tau_z$.

Evaluating the topological invariant in Eq. (4.2) again requires a $|\mathbf{k}| \to \infty$ regularization. One possibility is to add $\epsilon |\mathbf{k}|^2 \tau_x$ so that $|\mathbf{k}| \to \infty$ can be replaced by a single point. In this case the invariant can be determined by computing the winding degree of $\hat{\mathbf{d}}(\mathbf{k},\mathbf{r})$ on S^3. In the limit $\epsilon \to 0$ the $\mathbf{k}$ integral can be performed so that Eq. (4.2) can be expressed as the winding number of the phase of $\phi_1 + i\phi_2 = |\Delta| e^{i\varphi}$

$$n = \frac{1}{2\pi} \int_{S^1} d\varphi. \quad (4.7)$$

4. Hedgehogs in d = 3

In Ref. [53] we introduced a three dimensional model for Majorana bound states that can be interpreted as a theory of a vortex at the interface between a superconductor and a topological insulator. In the special case that the chemical potential is equal to zero, model has the same form as Eq. (4.6), except that now all of the vectors are three dimensional. In the topological insulator model we have $\tilde{\gamma} = (\gamma_1, \gamma_2, \gamma_3) = \mu_x \tau_z \vec{\sigma}$ and $\vec{\Gamma} = (\Gamma_1, \Gamma_2, \Gamma_3) = (\mu_z \tau_z, \tau_x, \tau_y)$. $\vec{\tau}$ and $\vec{\sigma}$ are defined as before while $\vec{\mu}$ describes a orbital degree of freedom. The chiral symmetry, $\Pi = \mu_y \tau_z$ is violated if a chemical potential term $\mu \tau_z$ is included.

Following the same steps that led to Eq. (4.7) the invariant Eq. (4.2) is given by the winding number of $\hat{\phi} = \vec{\phi}/|\vec{\phi}|$ on S^2

$$n = \frac{1}{4\pi} \int_{S^2} \hat{\phi} \cdot (d\hat{\phi} \times d\hat{\phi}). \quad (4.8)$$

B. Class D: Majorana bound states

1. Topological invariant

Point defects in class D are characterized by a $\mathbb{Z}_2$ topological invariant that determines the presence or absence of a Majorana bound state associated with the defect. These include the well-known end states in a 1D p-wave superconductor and vortex states in a 2D p_x+ip_y superconductor. In Ref. 53 we considered such zero modes in a three-dimensional BdG theory describing Majorana zero modes in topological insulator structures. Here we develop a unified description of all of these cases.

For a point defect in d dimensions, the Hamiltonian depends on d momentum variables and $D=d-1$ position variables. In Appendix D we show that the $\mathbb{Z}_2$ invariant is given by

$$\nu = \frac{2}{d!}\left(\frac{i}{2\pi}\right)^d \int_{T^d \times S^{d-1}} \mathcal{Q}_{2d-1} \bmod 2, \qquad (4.9)$$

where $\mathcal{Q}_{2d-1}$ is the Chern Simons form. The specific cases of interest are

$$\mathcal{Q}_1 = \text{Tr}[\mathcal{A}], \qquad (4.10)$$

$$\mathcal{Q}_3 = \text{Tr}\left[\mathcal{A}d\mathcal{A} + \frac{2}{3}\mathcal{A}^3\right], \qquad (4.11)$$

$$\mathcal{Q}_5 = \text{Tr}\left[\mathcal{A}(d\mathcal{A})^2 + \frac{3}{2}\mathcal{A}^3 d\mathcal{A} + \frac{3}{5}\mathcal{A}^5\right]. \qquad (4.12)$$

It is instructive to see that Eq. (4.9) reduces to Eq. (4.2) in the case in which a system also has particle-hole symmetry. In this case, as detailed in Appendix D it is possible to choose a gauge in which $\mathcal{A}=q^\dagger dq/2$, so that $\mathcal{Q}_{2d-1} \propto (q dq^\dagger)^{2d-1}$.

2. End states in a 1D superconductor

The simplest example of a point defect in a superconductor occurs in Kitaev's model[38] of a one-dimensional p-wave superconductor. This is described by a simple 1D tight binding model for spinless electrons, which includes a nearest-neighbor hopping term $tc_i^\dagger c_{i+1}+$H.c. and a nearest-neighbor p-wave pairing term $\Delta c_i c_{i+1}+$H.c.. The Bogoliubov de Gennes Hamiltonian can then be written as

$$\mathcal{H}(k) = (t\cos k - \mu)\tau_z + \Delta \sin k \tau_x. \qquad (4.13)$$

This model exhibits a weak pairing phase for $|\mu|<t$ and a strong pairing phase for $|\mu|>t$. The weak pairing phase will have zero-energy Majorana states at its ends.

The topological invariant in Eq. (4.9) can be easily evaluated. We find $\mathcal{A}=d\theta/2$, where θ is the polar angle of $\mathbf{d}(k) = (t\cos k-\mu, \Delta \sin k)$. It follows that for $|\mu|<t$, the topological invariant is $\nu=1 \bmod 2$.

3. Vortex in a 2D topological superconductor

In two dimensions, a Majorana bound state occurs at a vortex in a topological superconductor. This can be easily seen by considering the edge states of the topological superconductor in the presence of a hole.[39] Particle-hole symmetry requires that the quantized edge states come in pairs. When the flux is an odd multiple of $h/2e$, the edge states are quantized such that a zero mode is present. In this section we will evaluate the topological invariant in Eq. (4.9) associated with a loop surrounding the vortex.

We begin with the class D BdG Hamiltonian $\mathcal{H}_p^0(k_x, k_y)$ characterizing the topological superconductor when the superconducting phase is zero. We include the subscript p to denote the first Chern number that classifies the topological superconductor. We can then introduce a nonzero superconducting phase by a gauge transformation

$$\mathcal{H}_p(\mathbf{k}, \varphi) = e^{-i\varphi\tau_z/2}\mathcal{H}_p^0(\mathbf{k})e^{i\varphi\tau_z/2}, \qquad (4.14)$$

where τ_z operates in the Nambu particle-hole space. We now wish to evaluate Eq. (4.9) for this Hamiltonian when phase $\varphi(s)$ winds around a vortex. There is, however, a difficulty because the Chern Simons formula requires a gauge that is continuous throughout the entire base space $T^2 \times S^1$. The nonzero Chern number p characterizing $\mathcal{H}_p^0(\mathbf{k})$ is an obstruction to constructing such a gauge. A similar problem arose in Sec. III C 2, when we discussed a line dislocation in a weak topological superconductor. We can adapt the trick we used there to get around the present problem. We thus double the Hilbert space to include two copies of our Hamiltonian, one with Chern number p and one with Chern number $-p$

$$\widetilde{\mathcal{H}}^0(\mathbf{k}) = \begin{pmatrix} \mathcal{H}_p^0(\mathbf{k}) & 0 \\ 0 & \mathcal{H}_{-p}^0(\mathbf{k}) \end{pmatrix}. \qquad (4.15)$$

We then put the vortex in only the $+p$ component

$$\widetilde{\mathcal{H}}(\mathbf{k}, \varphi) = e^{-i\varphi q}\widetilde{\mathcal{H}}^0(\mathbf{k})e^{i\varphi q}, \qquad (4.16)$$

where

$$q = \frac{1+\tau_z}{2}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}. \qquad (4.17)$$

We added an extra phase factor by replacing τ_z by $1+\tau_z$ in order to make $e^{i\varphi q}$ periodic under $\varphi \to \varphi + 2\pi$.

Since the Chern number characterizing $\widetilde{\mathcal{H}}^0(\mathbf{k})$ is zero, there exists a continuous gauge

$$|\tilde{u}_i(\mathbf{k}, \varphi)\rangle = e^{i\varphi q}|\tilde{u}_i^0(\mathbf{k})\rangle, \qquad (4.18)$$

which allows us to evaluate the Chern-Simons integral. The Berry's connection $\widetilde{\mathcal{A}}_{ij}=\langle \tilde{u}_i | d\tilde{u}_j\rangle$ is given by

$$\widetilde{\mathcal{A}} = \widetilde{\mathcal{A}}^0 + iQd\varphi, \qquad (4.19)$$

where $\widetilde{\mathcal{A}}^0(\mathbf{k})$ is the connection describing $\widetilde{\mathcal{H}}^0(\mathbf{k})$ and $Q_{ij}(\mathbf{k}) = \langle \tilde{u}_i^0(\mathbf{k})|q|\tilde{u}_j^0(\mathbf{k})\rangle$. Inserting this into Eq. (4.11) and rearranging terms we find

$$\mathcal{Q}_3 = \text{Tr}[2Q\widetilde{\mathcal{F}}^0 - d(Q\widetilde{\mathcal{A}}^0)] \wedge d\varphi, \qquad (4.20)$$

where $\widetilde{\mathcal{F}}^0 = d\widetilde{\mathcal{A}}^0 + \widetilde{\mathcal{A}}^0 \wedge \widetilde{\mathcal{A}}^0$. Since the second term is a total derivative it can be discarded. For the first term there are two contributions from the 1 and the τ_z in Eq. (4.17). Upon inte-

grating over **k**, the τ_z term can be shown to vanish as a consequence of particle-hole symmetry. The 1 term simply projects out the Berry curvature of the original Hamiltonian $\mathcal{H}_p^0(\mathbf{k})$ so that

$$\mathcal{Q}_3 = \text{Tr}[\mathcal{F}^0] \wedge d\varphi. \qquad (4.21)$$

It follows from Eq. (4.9) that the $\mathbb{Z}_2$ invariant characterizing the vortex is

$$\nu = pm \bmod 2, \qquad (4.22)$$

where p is the Chern number characterizing the topological superconductor and m is the phase winding number associated with the vortex.

It is also instructive to consider this invariant in the context of the simple two-band model introduced by Read and Green.[39] This can be written as a simple tight-binding model

$$\mathcal{H}^0(k_x, k_y) = [t(\cos k_x + \cos k_y) - \mu]\tau_z + \Delta(\sin k_x \tau_x + \sin k_y \tau_y), \qquad (4.23)$$

where the superconducting order parameter Δ is real. As in Eq. (4.13), this model exhibits weak and strong pairing phases for $|\mu| < t$ and $|\mu| > t$. These are distinguished by the Chern invariant, which in turn is related to the winding number on S^2 of the unit vector $\hat{\mathbf{d}}(\mathbf{k})$, where $\vec{d}(\mathbf{k})$ are the coefficients of $\vec{\tau}$ in Eq. (4.23). A nonzero superconducting phase is again introduced by rotating about τ_z as in Eq. (4.14). Here we wish to show that in this two band model the $\mathbb{Z}_2$ invariant ν can be understood from a geometrical point of view.

The Z_2 invariant characterizing a vortex can be understood in terms of the topology of the maps $\hat{\mathbf{d}}(k_x, k_y, \phi)$. from $T^2 \times S^1$ to S^2. These maps were first classified by Pontrjagin[79] and have also appeared in other physical contexts.[53,80,81] Without losing generality, we can reduce the torus T^2 to a sphere S^2 so the mappings are $S^2 \times S^1 \to S^2$. When for fixed ϕ $\hat{\mathbf{d}}(k_x, k_y, \phi)$ has an S^2 winding number of $\pm p$, the topological classification is $\mathbb{Z}_{2p}$. In the case of interest, $p = 1$, so there are two classes.

This $\mathbb{Z}_2$ Pontrjagin invariant can be understood pictorially by considering inverse image paths in $(\mathbf{k}, \phi)$ space, which map to two specific points on S^2. These correspond to 1D curves in $S^2 \times S^1$. Figure 7 shows three examples of such curves. The inner sphere corresponds to $\phi = 0$ while the outer sphere corresponds to $\phi = 2\pi$. Since $p = 1$, for every point on S^2 the inverse image path is a single curve connecting the inner and outer spheres. The key point is to examine the linking properties of these curves. The $\mathbb{Z}_2$ invariant describes the number of twists in a pair of inverse image paths, which is 1 in (a), 2 in (b), and 0 in (c). The configuration in (b) can be continuously deformed into that in (c) by dragging the paths around the inner sphere. This can be verified by a simple demonstration using your belt. The twist in (a), however, cannot be undone. The number of twists thus defines the $\mathbb{Z}_2$ Pontrjagin invariant.

4. Superconductor heterostructures

Finally, in three dimensions, a nontrivial point defect can occur at a superconductor heterostructure. An example is a
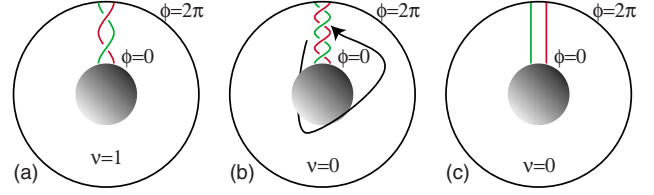

FIG. 7. (Color online) Visualization of the $\mathbb{Z}_2$ Pontrjagin invariant characterizing maps from $(\mathbf{k}, \phi) \in S^2 \times S^1$ to S^2 when the winding degree for each ϕ is 1. The inner sphere corresponds to **k** at $\phi = 0$ while the outer sphere is $\phi = 2\pi$. The lines depict inverse images of two specific points on S^2, which are lines connecting the inner and outer spheres. In (a) they have one twist, which cannot be eliminated. The double twist in (b) can be unwound by smoothly dragging the paths around the sphere to arrive at (c), which has no twist.

vortex in the superconducting state at the interface between a superconductor and a topological insulator. As shown in Ref. 53, this can be described by the simple Hamiltonian

$$\mathcal{H} = \upsilon \tau_z \mu_x \vec{\sigma} \cdot \mathbf{k} - \mu \tau_z + (m + \epsilon |\mathbf{k}|^2)\tau_z \mu_z + \Delta_1 \tau_x + \Delta_2 \tau_y. \qquad (4.24)$$

Here m is a mass which distinguishes a topological insulator from a trivial insulator, and $\Delta = \Delta_1 + i\Delta_2$ is a superconducting order parameter. For $\mu = 0$, this Hamiltonian has the form of the three dimensional version of Eq. (4.6) discussed in Sec. IV A 4, where the mass term is characterized by the vector $\vec{\phi} = (m, \Delta_1, \Delta_2)$. A vortex in Δ at the interface where m changes sign then corresponds to a hedgehog singularity in $\vec{\phi}$. From Eq. (4.3), it can be seen that the class BDI $\mathbb{Z}$ invariant is $n = 1$. This then establishes that the class D $\mathbb{Z}$ invariant is $\nu = 1$. The $\mathbb{Z}_2$ survives when a nonzero chemical potential reduces the symmetry from class BDI to class D.

C. Class DIII: Majorana doublets

Point defects in class DIII are characterized by a $\mathbb{Z}_2$ topological invariant. These are associated with zero modes, but unlike class D, the zero modes are required by Kramers theorem to be doubly degenerate. The zero modes thus form a Majorana doublet, which is equivalent to a single Dirac fermion.

In Table I, Class DIII, $\delta = 1$ is an entry that is similar to Class AII, $\delta = 2$. The $\mathbb{Z}_2$ for DIII invariant bears a resemblance to the invariant for AII, which is a generalization of the $\mathbb{Z}_2$ invariant characterizing the 2D quantum spin-Hall insulator. In Appendix B we will establish a formula that employs the same gauge constraint

$$w(\mathbf{k}, \mathbf{r}) = w_0, \qquad (4.25)$$

where w_0 is a constant independent of **k** and **r**. $w(\mathbf{k}, \mathbf{r})$ relates the time-reversed states at **k** and $-\mathbf{k}$ and is given by Eq. (3.23). *Provided* we choose a gauge that satisfies this constraint, the $\mathbb{Z}_2$ invariant is given by
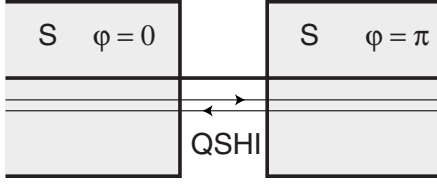
FIG. 8. A Josephson junction in proximity with the helical edge states of a quantum spin-Hall insulator. When the phase difference is π, there is a zero energy Majorana doublet at the junction.

$$\tilde{\nu} = \frac{1}{d!}\left(\frac{i}{2\pi}\right)^d \int_{T^d \times S^{d-1}} \mathcal{Q}_{2d-1} \bmod 2. \qquad (4.26)$$

This formula is almost identical to the formula for a point defect in class D but they differ by an important factor of two. Due to the combination of time-reversal and particle-hole symmetries the Chern-Simons integral in Eq. (4.26) is guaranteed to be an integer but the integer is not gauge invariant. When the time-reversal constraint is satisfied, the parity $\tilde{\nu}$ is gauge invariant. It then follows that the class D invariant in Eq. (4.9), $\nu=0$ mod 2.

In the special case $d=1$ there is a formula that does not rely on the gauge constraint, though it still requires a globally defined gauge. It is related to the similar "fixed-point" formula for the invariant for the 2D quantum spin-Hall insulator,[73] and has recently been employed by Qi, Hughes, and Zhang[82] to classify one-dimensional time-reversal invariant superconductors. In class DIII, it is possible to choose a basis in which the time-reversal and particle-hole operators are given by $\Theta=\tau_y K$ and $\Xi=\tau_x K$ so that the chiral operator is $\Pi=\tau_z$. In this basis, the Hamiltonian has the form (4.1), where $q(\mathbf{k},\mathbf{r}) \to q(k)$ satisfies $q(-k)=-q(k)^T$. Thus, $\mathrm{Pf}[q(k)]$ is defined for the time-reversal invariant points $k=0$ and $k=\pi$. $q(k)$ is related to $w(k)$ because in a particular gauge it is possible to choose $w(k)=q(k)/\sqrt{|\mathrm{Det}[q(k)]|}$. The $\mathbb{Z}_2$ invariant is then given by

$$(-1)^{\tilde{\nu}} = \frac{\mathrm{Pf}[q(\pi)]}{\mathrm{Pf}[q(0)]} \frac{\sqrt{\mathrm{Det}[q(0)]}}{\sqrt{\mathrm{Det}[q(\pi)]}}, \qquad (4.27)$$

where the branch $\sqrt{\mathrm{Det}[q(k)]}$ is chosen continuously between $k=0$ and $k=\pi$. The equivalence of Eqs. (4.26) and (4.27) for $d=1$ is demonstrated in Appendix A. Unlike Eq. (4.26), however, the fixed-point formula (4.27) does not have a natural generalization for $d>1$.

Majorana doublets can occur at topological defects in time-reversal invariant topological superconductors, or in Helium 3B. Here we consider a different configuration at a Josephson junction at the edge of a quantum spin-Hall insulator (Fig. 8). When the phase difference across the Josephson junction is π, it was shown in Refs. 38 and 83 that there is a level crossing in the Andreev bound states at the junction. This corresponds precisely to a Majorana doublet.

This can be described by the simple continuum 1D theory introduced in Ref. 83

$$\mathcal{H} = vk\tau_z\sigma_z + \Delta_1\tau_x. \qquad (4.28)$$

Here σ_z describes the spin of the quantum spin-Hall edge state and Δ_1 is the real superconducting order parameter. This model has particle-hole symmetry $\Xi=\sigma_y\tau_y K$ and time-reversal symmetry $\Theta=i\sigma_y K$ and is in class DIII. A π junction corresponds to a domain wall where Δ_1 changes sign. Following appendix C, it is straightforward to see that this will involve a degenerate pair of zero modes indexed by the spin σ_z and chirality τ_y constrained by $\tau_y\sigma_z=-1$.

The Hamiltonian (4.28) should be viewed as a low-energy theory describing the edge of a 2D quantum spin-Hall insulator. Nonetheless, we may describe a domain wall where Δ_1 changes sign using an effective one dimensional theory by introducing a regularization replacing Δ_1 by $\Delta_1+\epsilon k^2$. This regularization will not affect the topological structure of a domain wall, where Δ_1 changes sign. A topologically equivalent lattice version of the theory then has the form

$$\mathcal{H} = t\sin k\tau_z\sigma_z + [\Delta_1 + u(1-\cos k)]\tau_x. \qquad (4.29)$$

where we assume $|\Delta_1|<2u$.

The topological invariant can be evaluated using either Eq. (4.26) or (4.27). To use Eq. (4.26), note that Eq. (4.29) has exactly the same form as two copies (distinguished by $\sigma_z=\pm 1$) of Eq. (4.13). The evaluation of Eq. (4.26) then proceeds along the same lines. It is straightforward to check that in a basis where the time-reversal constraint in Eq. (4.25) is satisfied (this fixes the relative phases of the $\sigma_z=\pm 1$ states), $\mathcal{A}=d\theta$, where θ is the polar angle of $\mathbf{d}(k)=[t\sin k, \Delta_1+u(1-\cos k)]$. It follows that a defect where Δ_1 changes sign has $\tilde{\nu}=1$.

To use Eq. (4.27), we transform to a basis in which $\Theta=\tau_y K$, $\Xi=\tau_x K$, and $\Pi=\tau_z$. This is accomplished by the unitary transformation $U=\exp[i(\pi/4)\sigma_y\tau_z]\exp[i(\pi/4)\tau_x]$. Then, $\mathcal{H}$ has the form of Eq. (3.13) with $q(k)=-i\{t\sin k\sigma_z+[\Delta_1+u(1-\cos k)]\sigma_y\}$. It follows that $\det[q(k)]$ is real and positive for all k. Moreover, $\mathrm{Pf}[q(0)]/\sqrt{\det[q(0)]}=\mathrm{sgn}[\Delta_1]$ while $\mathrm{Pf}[q(\pi)]/\sqrt{\det[q(\pi)]}=1$. Again, a defect where Δ_1 changes sign has $\tilde{\nu}=1$.

V. ADIABATIC PUMPS

In this section we will consider time-dependent Hamiltonians $\mathcal{H}(\mathbf{k},\mathbf{r},t)$, where in additional to having adiabatic spatial variation $\mathbf{r}$ there is a cyclic adiabatic temporal variation parameterized by t. We will focus on pointlike spatial defects, in which the dimensions of $\mathbf{k}$ and $\mathbf{r}$ are related by $d-D=1$.

Adiabatic cycles in which $\mathcal{H}(\mathbf{k},\mathbf{r},t=T)=\mathcal{H}(\mathbf{k},\mathbf{r},t=0)$ can be classified topologically by considering t to be an additional "spacelike" variable, defining $\tilde{D}=D+1$. Such cycles will be classified by the $\delta=0$ column of Table I. Topologically nontrivial cycles correspond to adiabatic pumps. Table IV shows the symmetry classes which host nontrivial pumping cycles, along with the character of the adiabatic pump. There are two general cases. Classes A, AI, and AII define a charge pump, where after one cycle an integer number of charges is transported toward or away from the point defect.

TABLE IV. Symmetry classes that support nontrivial charge or fermion parity pumping cycles.

Symmetry	Topological classes	Adiabatic Pump
A	$\mathbb{Z}$	Charge
AI	$\mathbb{Z}$	Charge
BDI	$\mathbb{Z}_2$	Fermion parity
D	$\mathbb{Z}_2$	Fermion parity
AII	$2\mathbb{Z}$	Charge Kramers doublet

Classes BDI and D define a fermion parity pump. We will discuss these two cases separately.

We note in passing that the $\delta=0$ column of Table I also applies to topological *textures*, for which $d=D$. For example, a spatially dependent three-dimensional band structure $\mathcal{H}(\mathbf{k},\mathbf{r})$ can have topological textures analogous to Skyrmions in a 2D magnet. Such textures have recently been analyzed by Ran, Hosur, and Vishwanath[84] for the case of class D, where they showed that the $\mathbb{Z}_2$ invariant characterizing the texture corresponds to the fermion parity associated with the texture. Thus, nontrivial textures are fermions.

A. Classes A, AI, and AII: Thouless charge pumps

The integer topological invariant characterizing a pumping cycle in class A is simply the Chern number characterizing the Hamiltonian $\mathcal{H}(\mathbf{k},\mathbf{r},t)$.[60,61] Imposing time-reversal symmetry has only a minor effect on this. For $\Theta^2=-1$ (Class AII), an odd Chern number violates time-reversal symmetry so that only even Chern numbers are allowed. This means that the pumping cycle can only pump Kramers pairs of electrons. For $\Theta^2=+1$ (Class AI) all Chern numbers are consistent with time-reversal symmetry.

The simplest charge pump is the 1D model introduced by Thouless.[60] A continuum version of this model can be written in the form

$$\mathcal{H}(k,t)=vk\sigma_z+[m_1(t)+\epsilon k^2]\sigma_x+m_2(t)\sigma_y. \quad (5.1)$$

When the masses undergo a cycle such that the phase of m_1+im_2 a single electron is transmitted down the wire. In this case, $\mathcal{H}(k,t)$ has a nonzero first Chern number. The change in the charge associated with a point in a 1D system is given by the difference in the Chern numbers associated with either side of the point. Thus, after a cycle a charge e accumulates at the end of a Thouless pump.

A two-dimensional version of the charge pump can be developed based on Laughlin's argument[85] for the integer quantum-Hall effect. Consider a 2D $\nu=1$ integer quantum-Hall state and change the magnetic flux threading a hole from 0 to h/e. In the process, a charge e is pumped to the edge states surrounding the hole. This pumping process can be characterized by the second Chern number characterizing the 2D Hamiltonian $\mathcal{H}(k_x,k_y,\theta,t)$, where θ parameterizes a circle surrounding the hole. A similar pump in 3D can be considered and is characterized by the third Chern number.

B. Class D, BDI: Fermion parity pump

Adiabatic cycles of point defects in class D and BDI are characterized by a $\mathbb{Z}_2$ topological invariant. In this section we will argue that a nontrivial pumping cycle transfers a unit of fermion parity to the point defect. This is intimately related to the Ising non-Abelian statistics associated with defects supporting Majorana bound states.

Like the point defect in class DIII ($\delta=1$), the temporal pump ($\delta=0$) in class D occupies an entry in Table I similar to the line defect ($\delta=2$) in class AII so we expect a formula that is similar to the formula for the 2D quantum spin-Hall insulator. This is indeed the case, though the situation is slightly more complicated. The Hamiltonian $\mathcal{H}(\mathbf{k},\mathbf{r},t)$ is defined on a base space $T^d\times S^{d-1}\times S^1$. In Appendix F we will show that the invariant can be written in a form that resembles Eq. (3.25)

$$\nu=\frac{i^d}{d!(2\pi)^d}\left[\int_{\mathcal{T}_{1/2}}\mathrm{Tr}(\mathcal{F}^d)-\oint_{\partial\mathcal{T}_{1/2}}\mathcal{Q}_{2d-1}\right]\mathrm{mod}\ 2, \quad (5.2)$$

where $\mathcal{T}_{1/2}$ is half of the base manifold, say, $k_1\in[0,\pi]$, and the Chern-Simons form $\mathcal{Q}_{2d-1}$ is generated by a continuous valence frame $u_v(\mathbf{k},\mathbf{r},t)|_{k_1=0,\pi}$ that obeys certain particle-hole gauge constraint. This is more subtle than the time-reversal gauge condition in Eq. (3.24) for line defects in AII and point defects in DIII. Unlike Eq. (3.24), we do not have a computational way of checking whether or not a given frame satisfies the constraint. Nevertheless, it can be defined, and in certain simple examples, the particle-hole constraint is automatically satisfied.

The origin of the difficulty is that unlike time-reversal symmetry, particle-hole symmetry connects the conduction and valence bands. The gauge constraint therefore involves both. Valence and conduction frames can be combined to form a unitary matrix

$$G_{\mathbf{k},\mathbf{r},t}=\begin{pmatrix} | & | \\ u_v(\mathbf{k},\mathbf{r},t) & u_c(\mathbf{k},\mathbf{r},t) \\ | & | \end{pmatrix}\in U(2n). \quad (5.3)$$

The orthogonality of conduction and valence band states implies that

$$G_{\mathbf{k},\mathbf{r},t}^\dagger \Xi G_{-\mathbf{k},\mathbf{r},t}=0. \quad (5.4)$$

In general, we call a frame $G:\partial\mathcal{T}_{1/2}\to U(2n)$ particle-hole trivial if it can continuously be deformed to a constant *while satisfying Eq. (5.4) throughout the deformation*. The Chern Simons term in Eq. (5.2) requires a gauge that is built from the valence-band part of a particle-hole trivial frame.

Though the subtlety of the gauge condition makes a general computation of the invariant difficult, it is possible to understand the invariant in the context of specific models. Consider, a theory based on a point defect in the d-dimensional version of Eq. (4.6)

$$\mathcal{H}(\mathbf{k},\mathbf{r},t) = v\,\tilde{\gamma}\cdot\mathbf{k} + \vec{\Gamma}\cdot\vec{\phi}(\mathbf{r},t). \qquad (5.5)$$

Here $\vec{\Gamma}$ and $\tilde{\gamma}$ are $2^d \times 2^d$ Dirac matrices, and we suppose that for fixed t, the d-dimensional mass vector $\vec{\phi}(\mathbf{r},t)$ has a point topological defect at $\mathbf{r}_0(t)$. If Ref. 53 we argued that adiabatic cycles for such point defects are classified by a Pontrjagin invariant similar to that discussed in Sec. IV B 3. This may also be understood in terms of the rotation of the "orientation" of the defect. Near the defect, suppose $\vec{\phi}(\mathbf{r},t)=O(t)[\mathbf{r}-\mathbf{r}_0(t)]$, where $O(t)$ is a time-dependent $O(d)$ rotation. In the course of the cycle, the orientation of the topological defect, characterized by $O(t)$ goes through a cycle. Since for $d \geq 3$, $\pi_1[O(d)]=\mathbb{Z}_2$, there are two classes of cycles. As shown in Ref. 53, the nontrivial cycle, which corresponds to a 2π rotation changes the sign of the Majorana fermion wave function associated with the topological defect. We will argue below that this corresponds to a change in the local fermion parity in the vicinity of the defect. For $d=2$, $\pi_1[O(2)]=\mathbb{Z}$. However, the change in the sign of the Majorana bound state is given by the parity of the $O(2)$ winding number. In theories with more bands, it is only this parity that is topologically robust.

In $d=1$, the single Γ matrix in the two-band model does not allow for continuous rotations. Consider instead Kitaev's model[38] for a 1D topological superconductor with at time-dependent phase

$$\mathcal{H}(k,t) = (t\cos k - \mu)\tau_z + \Delta_1(t)\sin k\,\tau_x + \Delta_2(t)\sin k\,\tau_y. \qquad (5.6)$$

In this case it is possible to apply the formula (5.2) because on the boundary $\partial\mathcal{T}$, which is $k=0$ or $k=\pi$ the Hamiltonian is independent of t, so that the gauge condition in Eq. (5.4) is automatically satisfied. Moreover, the second term in Eq. (5.2) involving the Chern Simons integral is equal to zero so that the invariant is simply the integral of $\mathcal{F}(x,t)$ over $\mathcal{T}_{1/2}$. It is straightforward to check that this gives $\nu=1$.

In order to see why this corresponds to a pump for fermion parity, suppose a topological superconductor is broken in two places, as shown in Fig. 9. At the ends where the superconductor is cut there will be Majorana bound states. The pair of bound states associated with each cut defines two quantum states which differ by the *parity* of the number of electrons. If the two ends are weakly coupled by electron tunneling then the pair of states will split. Now consider advancing the phase of the central superconductor by 2π. As shown in Refs. 38 and 83, the states interchange as depicted in Fig. 9. The level crossing that occurs at π phase difference is protected by the conservation of fermion parity. Thus, at the end of the cycle, one unit of fermion parity has been transmitted from one circled region to the other. The pumping of fermion parity also applies to adiabatic cycles of point defects in higher dimensions and is deeply connected with the Ising non-Abelian statistics associated with those defects.[53]

VI. CONCLUSION

In this paper we developed a unified framework for classifying topological defects in insulators and superconductors
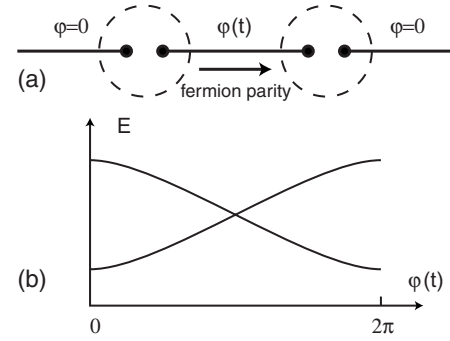

FIG. 9. A one-dimensional fermion parity pump based on a 1D topological superconductor, which has Majorana states at its ends. When the phase of the central superconductor is advanced by 2π the fermion parity associated with the pairs of Majorana states inside each circle changes. Thus fermion parity has been pumped from one circle to the other. (b) shows the evolution of the energy levels associated with a weakly coupled pair of Majorana states as a function of phase. The level crossing at $\phi=\pi$ is protected by the local conservation of fermion parity.

by considering Bloch/BdG Hamiltonians that vary adiabatically with spatial (and/or temporal) parameters. This led to a generalization of the bulk-boundary correspondence, which identifies protected gapless fermion excitations with topological invariants characterizing the defect. This leads to a number of additional questions to be addressed in future work.

The generalized bulk-boundary correspondence has the flavor of a mathematical index theorem, which relates an analytic index that characterizes the zero modes of a system to a topological index. It would be interesting to see a more general formulation of this relation[86,87] that applies to the classes without chiral symmetry that have $\mathbb{Z}_2$ invariants and goes beyond the adiabatic approximation we used in this paper. Though the structure of the gapless modes associated with defects make it clear that such states are robust in the presence of disorder and interactions, it would be desirable to have a more general formulation of the topological invariants characterizing a defect that can be applied to interacting and/or disordered problems.

An important lesson we have learned is that topologically protected modes can occur in a context somewhat more general than simply boundary modes. This expands the possibilities for engineering these states in physical systems. It is thus an important future direction to explore the possibilities for heterostructures that realize topologically protected modes. The simplest version of this would be to engineer protected chiral fermion modes using a magnetic topological insulator. The perfect electrical transport in such states could have far reaching implications at both the fundamental and practical level. In addition, it is worth considering the expanded possibilities for realizing Majorana bound states in superconductor heterostructures, which could have implications for quantum computing.

Finally, it will be interesting to generalize these topological considerations to describe inherently correlated states, such as the Laughlin state. Could a *fractional* quantum-Hall edge state arise as a topological line defect in a 3D system?
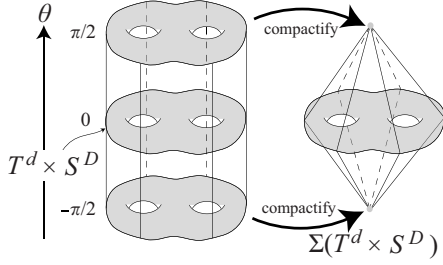
FIG. 10. Suspension $\Sigma(T^d \times S^D)$. The top and bottom of the cylinder $\Sigma(T^d \times S^D) \times [-\pi/2, \pi/2]$ are identified to two points.

Understanding the topological invariants that would characterize such a defect would lead to a deeper understanding of topological states of matter.

APPENDIX A: PERIODICITY IN SYMMETRY AND DIMENSION

In this appendix we will establish the relations in Eqs. (2.5) and (2.6) between the K groups in different position-momentum dimensions (D, d) and different symmetry classes s. We will do so by starting with an arbitrary Hamiltonian in $K_{\mathbb{F}}(s; D, d)$ and then explicitly constructing new Hamiltonians in one higher position or momentum dimension, which have a symmetry either added or removed. The new Hamiltonians will then belong to $K_{\mathbb{F}}(s+1; D, d+1)$ or $K_{\mathbb{F}}(s-1; D+1, d)$. The first step is to identify the mappings and show they preserve the group structure. This defines group homomorphisms relating the K groups. The next step is to show they are *isomorphisms* by showing that the maps have an inverse, up to homotopic equivalence.

1. Hamiltonian mappings

There are two classes of mappings: those that add symmetries and those that remove symmetries. These need to be considered separately.

We consider first the symmetry removing mappings that send a Hamiltonian $\mathcal{H}_c$ with chiral symmetry to a Hamiltonian $\mathcal{H}_{nc}$ without chiral symmetry. Suppose $\{\mathcal{H}_c(\mathbf{k}, \mathbf{r}), \Pi\} = 0$, where Π is the chiral operator. Then define

$$\mathcal{H}_{nc}(\mathbf{k}, \mathbf{r}, \theta) = \cos\theta \mathcal{H}_c(\mathbf{k}, \mathbf{r}) + \sin\theta\Pi \quad \text{(A1)}$$

for $-\pi/2 \leq \theta \leq \pi/2$. This has the property that at $\theta = \pm\pi/2$ the new Hamiltonian is $\pm\Pi$, independent of $\mathbf{k}$ and $\mathbf{r}$. Thus, at each of these points we may consider the base space $T^d \times S^D$ defined by $\mathbf{k}$ and $\mathbf{r}$ to be contracted to a point. The new Hamiltonian is then defined on the *suspension* $\Sigma(T^d \times S^D)$ of the original base space (see Fig. 10). If we treat the original
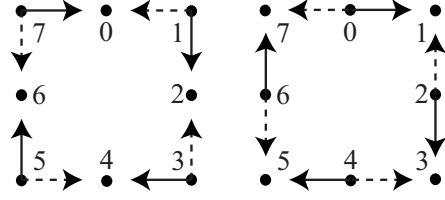

FIG. 11. Hamiltonian mappings in Eqs. (A1) and (A3) are drawn on the left and right clocks, respectively. Solid (dotted) arrows represent addition of one momentum (spatial) dimension.

base space as a $d+D$-dimensional sphere, then the suspension is a $d+D+1$-dimensional sphere.

Without loss of generality we assume $\mathcal{H}_c$ is flattened so that $\mathcal{H}_c^2 = 1$. Since $\{\mathcal{H}_c, \Pi\} = 0$ it follows that $\mathcal{H}_{nc}^2 = 1$ as well. The second term in Eq. (A1) violates the chiral symmetry. Thus, if $\mathcal{H}_c$ belongs to the complex class AIII (with no antiunitary symmetries), then $\mathcal{H}_{nc}$ belongs to class A. Equation (A1) thus provides a mapping from class AIII to class A.

For the real classes, which have antiunitary symmetries, the second term will violate either particle-hole symmetry or time-reversal symmetry, depending on whether θ is a momentum or position-type variable (odd or even under Θ and Ξ). This will lead to a new nonchiral symmetry class related to the original class by either a clockwise or counterclockwise turn on the symmetry clock (Fig. 11). To determine which it is, note that if we require $[\Theta, \Xi] = 0$ then $(\Theta\Xi)^2 = \Theta^2\Xi^2 = (-1)^{(s-1)/2}$. The unitary chiral symmetry operator (satisfying $\Pi^2 = 1$) can then be written

$$\Pi = i^{(s-1)/2}\Theta\Xi. \quad \text{(A2)}$$

It follows that if θ is momentumlike, then time-reversal symmetry is violated when $s = 1 \bmod 4$, while particle hole is violated when $s = 3 \bmod 4$. This corresponds to corresponds to a clockwise rotation on the symmetry clock, $s \to s+1$. If θ is position like then $s \to s-1$.

We next build a chiral Hamiltonian from a nonchiral one by adding a symmetry. This is accomplished by doubling the number of bands in a manner similar to the doubling employed in the Bogoliubov de Gennes description of a superconductor. We thus write

$$\mathcal{H}_c(\mathbf{k}, \mathbf{r}, \theta) = \cos\theta \mathcal{H}_{nc}(\mathbf{k}, \mathbf{r}) \otimes \tau_z + \sin\theta\mathbb{1} \otimes \tau_a, \quad \text{(A3)}$$

where $a = x$ or y. Here $\vec{\tau}$ are Pauli matrices that act on the doubled degree of freedom. As in Eqs. (A1), Eq. (A3) gives a new Hamiltonian defined on a base space that is the suspension of the original base space. If $\mathcal{H}_{nc}^2 = 1$ it follows that $\mathcal{H}_c^2 = 1$ so the energy gap is preserved. It is also clear that the new Hamiltonian has a chiral symmetry because it anticommutes with $\Pi = i\tau_z\tau_a$. Thus, if $\mathcal{H}_{nc}$ is in class A, then $\mathcal{H}_c$ is in class AIII.

For the real-symmetry classes $a = x$ or y must be chosen so that the second term in Eq. (A3) preserves the original antiunitary symmetry of $\mathcal{H}_{nc}$. This depends on the original antiunitary symmetry and whether θ is chosen to be a momentum or a position variable. For example, if $\mathcal{H}_{nc}$ has time-reversal symmetry, Θ, and θ is a momentum (position) variable, then we require $a = y$ ($a = x$). In this case, $\mathcal{H}_c$ has the

additional particle-hole symmetry $\Xi = \tau_x \Theta$ ($\Xi = i\tau_y \Theta$) that satisfies $\Xi^2 = \Theta^2$ ($\Xi^2 = -\Theta^2$). A similar analysis when $\mathcal{H}_{nc}$ has particle-hole symmetry allows us to conclude that the symmetry class of $\mathcal{H}_c$ is given by a clockwise rotation on the symmetry clock, $s \to s+1$, when θ is a momentum variable. When θ is a position variable, $s \to s-1$ gives a counterclockwise rotation.

Equations (A1) and (A3) map a Hamiltonian into a new Hamiltonian in a different dimension and different symmetry class. It is clear that two Hamiltonians that are topologically equivalent will be mapped to topologically equivalent Hamiltonians since the mapping can be done continuously on a smooth interpolation between the original Hamiltonians. Thus, Eqs. (A1) and (A3) define a mapping between equivalence classes of Hamiltonians. Moreover, since the direct sum of two Hamiltonians is mapped to the direct sum of the new Hamiltonians, the group property of the equivalence classes is preserved. Equations (A1) and (A3) thus define a K-group homomorphism

$$K_{\mathbb{F}}(s;D,d) \to K_{\mathbb{F}}(s+1;D,d+1), \qquad (A4)$$

$$K_{\mathbb{F}}(s;D,d) \to K_{\mathbb{F}}(s-1;D+1,d) \qquad (A5)$$

for $\mathbb{F} = \mathbb{R}, \mathbb{C}$.

2. Invertibility

In order to establish that Eqs. (A4) and (A5) are isomorphisms we need to show that there exists an inverse. This is *not* true of the Hamiltonian mappings. A general Hamiltonian cannot be built from a lower dimensional Hamiltonian using Eqs. (A1) and (A3). However, we will argue that it is possible to continuously deform any Hamiltonian into the form given by Eq. (A1) or (A3). Thus, the mappings between equivalence classes have an inverse. To show this we will use a mathematical method borrowed from Morse theory.[67]

Without loss of generality we again consider *flattened* Hamiltonians having equal number of conduction and valence bands with energies ± 1. Consider $\mathcal{H}(\mathbf{k}, \mathbf{r}, \theta)$, where $\theta \in [-\pi/2, \pi/2]$ is either a position or momentum variable and $\mathcal{H}$ is independent of $\mathbf{k}$ and $\mathbf{r}$ at $\theta = \pm \pi/2$. We wish to show that $\mathcal{H}(\mathbf{k}, \mathbf{r}, \theta)$ can be continuously deformed into the form (A1) or (A3). To do so we define an artificial "action"

$$S[\mathcal{H}(\mathbf{k}, \mathbf{r}, \theta)] = \int d\theta d^d\mathbf{k} d^D\mathbf{r} \, \text{Tr}(\partial_\theta \mathcal{H} \partial_\theta \mathcal{H}). \qquad (A6)$$

S can be interpreted as a "height" function in the space of gapped symmetry preserving Hamiltonians. Given any Hamiltonian there is always a downhill direction. These downhill vectors can then be integrated into a deformation trajectory. Since the action is positive definite, it is bounded below. The deformation trajectory must end at a Hamiltonian that locally minimizes the action.

Under the flatness constraint $\mathcal{H}^2 = 1$, minimal Hamiltonians satisfy the Euler-Lagrange equation

$$\partial_\theta^2 \mathcal{H} + \mathcal{H} = 0. \qquad (A7)$$

The solutions must be a linear combination of $\sin \theta$ and $\cos \theta$. The coefficient of $\sin \theta$ must be constant because the

base space is compactified to points at $\theta = \pm \pi/2$. A minimal Hamiltonian thus has the form

$$\mathcal{H}(\mathbf{k}, \mathbf{r}, \theta) = \cos \theta \mathcal{H}_1(\mathbf{k}, \mathbf{r}) + \sin \theta \mathcal{H}_0. \qquad (A8)$$

The constraint $\mathcal{H}(\mathbf{k}, \mathbf{r}, \theta)^2 = 1$ requires

$$\mathcal{H}_0^2 = \mathcal{H}_1(\mathbf{k}, \mathbf{r})^2 = 1, \quad \{\mathcal{H}_0, \mathcal{H}_1(\mathbf{k}, \mathbf{r})\} = 0. \qquad (A9)$$

If $\mathcal{H}(\mathbf{k}, \mathbf{r}, \theta)$ is nonchiral, then Eq. (A8) is already in the form of Eq. (A1) with $\Pi = \mathcal{H}_0$ and $\mathcal{H}_c(\mathbf{k}, \mathbf{r}) = \mathcal{H}_1(\mathbf{k}, \mathbf{r})$. $\mathcal{H}_1$ automatically has chiral symmetry due to Eq. (A9). This shows that Eqs. (A4) and (A5) are invertible when s is odd.

If $\mathcal{H}(\mathbf{k}, \mathbf{r}, \theta)$ is chiral, then both $\mathcal{H}_0$ and $\mathcal{H}_1(\mathbf{k}, \mathbf{r})$ anticommute with the chiral symmetry operator Π. Rename $\mathcal{H}_0 = \tau_a$ and $\Pi = i\tau_z \tau_a$, where $a = x$ ($a = y$) when θ is a position (momentum) variable. It follows that $\{\mathcal{H}_1, \tau_x\} = \{\mathcal{H}_1, \tau_y\} = 0$ so we can write

$$\mathcal{H}_1(\mathbf{k}, \mathbf{r}) = h(\mathbf{k}, \mathbf{r}) \otimes \tau_z. \qquad (A10)$$

Equation (A8) thus takes the form of Eq. (A3) with $\mathcal{H}_{nc} = h$. Since τ_z anticommutes with either Θ or Ξ, $h(\mathbf{k}, \mathbf{r})$ carries exactly one antiunitary symmetry and is therefore nonchiral. This shows that Eqs. (A4) and (A5) are invertible when s is even.

APPENDIX B: REPRESENTATIVE HAMILTONIANS AND CLASSIFICATION BY WINDING NUMBERS

In this appendix we construct representative Hamiltonians for each of the symmetry classes that are built as linear combinations of Clifford algebra generators that can be represented as anticommuting Dirac matrices. This allows us to relate the integer topological invariants, corresponding to the $\mathbb{Z}$ and $2\mathbb{Z}$ entries in Table I, to the winding degree in maps between spheres. Similar construction for defectless bulk Hamiltonians can be found in Ref. 68 by Ryu, *et al.*. In general, Hamiltonians do not have this specific form. However, since each topological class of Hamiltonians includes representatives of this form, it is always possible to smoothly deform $\mathcal{H}(\mathbf{k}, \mathbf{r})$ into this form.

The simplest example of this approach is the familiar case of a two-dimensional Hamiltonian with no symmetries (class A). A topologically nontrivial Hamiltonian can be represented as a 2×2 matrix that can be expressed in terms of Pauli matrices as $\mathcal{H}(\mathbf{k}) = \mathbf{h}(\mathbf{k}) \cdot \vec{\sigma}$. The Hamiltonian can then be associated with a unit vector $\hat{\mathbf{d}}(\mathbf{k}) = \mathbf{h}(\mathbf{k})/|\mathbf{h}(\mathbf{k})| \in S^2$. It is then well known that the Chern number characterizing $\mathcal{H}(\mathbf{k})$ in two dimensions is related to the *degree*, or winding number, of the mapping from $\mathbf{k}$ to S^2. This approach also applies to higher Chern numbers characterizing Hamiltonians in even dimensions $d = 2n$. In this case, a Hamiltonian that is a combination of $2n+1$ $2^n \times 2^n$ Dirac matrices, and can be associated with a unit vector $\hat{\mathbf{d}} \in S^{2n}$.

For the complex chiral class AIII, the $U(n)$ winding number characterizing a family of Hamiltonians can similarly be expressed as a winding number on spheres. For example, in $d = 1$, a chiral Hamiltonian can be written $\mathcal{H}(k) = h_x(k)\sigma_x + h_y(k)\sigma_y$ (so $\{\mathcal{H}, \sigma_z\} = 0$), and is characterized by $\hat{\mathbf{d}}(k) \in S^1$.

TABLE V. Examples of Dirac matrices for $(p,q)=(s,0)$.

Classes		Dirac matrices					Symmetry operators		
s	AZ	Γ_0	$\vec{\gamma}$				Θ	Ξ	Π
0	AI	$\mathbb{1}$					K		
1	BDI	σ_z	σ_y				K	$\sigma_x K$	σ_x
2	D	σ_z	σ_y	σ_x				$\sigma_x K$	
3	DIII	$\tau_z \sigma_z$	$\tau_z \sigma_y$	$\tau_z \sigma_x$	τ_x		$i\tau_y \sigma_x K$	$\sigma_x K$	τ_y
4	AII	$\tau_z \sigma_z$	$\tau_z \sigma_y$	$\tau_z \sigma_x$	τ_x	τ_y	$i\tau_y \sigma_x K$		

The integer topological invariant can then be expressed by the winding number of $\hat{\mathbf{d}}(k)$. Similar considerations apply to the integer invariants for chiral Hamiltonians in higher odd dimensions.

For the real symmetry classes we introduce "position-type" Dirac matrices Γ_μ and "momentum-type" Dirac matrices γ_i. These satisfy $\{\Gamma_\mu, \Gamma_\nu\}=2\delta_{\mu\nu}$, $\{\gamma_i, \gamma_j\}=2\delta_{ij}$, and $\{\Gamma_\mu, \gamma_j\}=0$, and are distinguished by their symmetry under antiunitary symmetries. If there is time-reversal symmetry we require

$$[\Gamma_\mu, \Theta] = \{\gamma_i, \Theta\} = 0 \tag{B1}$$

while with particle-hole symmetry

$$\{\Gamma_\mu, \Xi\} = [\gamma_i, \Xi] = 0. \tag{B2}$$

For a Hamiltonian that is a combination of p momentumlike matrices $\gamma_{1,\dots,p}$ and $q+1$ positionlike matrices $\Gamma_{0,\dots,q}$

$$\mathcal{H}(\mathbf{k}, \mathbf{r}) = \mathbf{R}(\mathbf{k}, \mathbf{r}) \cdot \vec{\Gamma} + \mathbf{K}(\mathbf{k}, \mathbf{r}) \cdot \vec{\gamma} \tag{B3}$$

the coefficients must satisfy the involution

$$\mathbf{R}(-\mathbf{k}, \mathbf{r}) = \mathbf{R}(\mathbf{k}, \mathbf{r}), \tag{B4}$$

$$\mathbf{K}(-\mathbf{k}, \mathbf{r}) = -\mathbf{K}(\mathbf{k}, \mathbf{r}). \tag{B5}$$

This can be characterized by a unit vector

$$\hat{\mathbf{d}}(\mathbf{k}, \mathbf{r}) = \frac{(\mathbf{K}, \mathbf{R})}{\sqrt{|\mathbf{K}|^2 + |\mathbf{R}|^2}} \in S^{p+q}, \tag{B6}$$

where S^{p+q} is a $(p+q)$ sphere in which p of the dimensions are odd under the involution in Eq. (B5).

The symmetry class s of $\mathcal{H}(\mathbf{k}, \mathbf{r})$ is related to the indices (p,q) characterizing the numbers of Dirac matrices by

$$p - q = s \bmod 8. \tag{B7}$$

To see this, start with a Hamiltonian $\mathcal{H}_0 = R_0(\mathbf{k}, \mathbf{r})\Gamma_0$ that involves a single 1×1 position like "Dirac matrix" $\Gamma_0 = \mathbb{1}$ so $(p,q) = (0,0)$. This clearly has time-reversal symmetry, with $\Theta = K$, and corresponds to class AI with $s=0$. Next, generate Hamiltonians $\mathcal{H}_s$ with different symmetries s by using the Hamiltonian mappings introduced in Appendix A. Both the mappings in Eqs. (A1) and (A3) define a new Clifford algebra with one extra generator that is either position or momentum type. The mappings that correspond to clockwise rotations on the symmetry clock $(s \to s+1)$ introduce an additional positionlike generator $(p \to p+1)$ while the map-

pings that correspond to counterclockwise rotations $(s \to s-1)$ introduce an additional momentumlike generator $(q \to q+1)$. Equation (B7) follows because this procedure can be repeated to generate Hamiltonians with any indices (p,q). Some examples are listed in Table V.

The integer topological invariants in Table I (which occur when $s-\delta$ is even) can be related to the winding degree of the maps $\hat{\mathbf{d}}: S^{D+d} \to S^{p+q}$. This can be nonzero when the spheres have the same total dimensions. In light of Eq. (B7), (p,q) can always be chosen so that $d+D=p+q$. The antiunitary symmetries impose constraints on the possible values of these winding numbers, which depend on the relation between $\delta = d-D$ and $s = p-q$.

The involutions on S^{d+D} and S^{p+q} have opposite orientations when $\delta - s \equiv 2$ or $6 \bmod 8$, and therefore an involution preserving map $S^{d+D} \to S^{p+q}$ can have nonzero winding degree only when $\delta - s \equiv 0$ or $4 \bmod 8$. Symmetry gives a further constraint on the latter case. Consider a sphere map $S^2_{\theta,\phi} \to S^2_{\vartheta,\varphi}$, where the involutions on the spheres send $(\theta,\phi) \mapsto (\theta, \phi+\pi)$ and $(\vartheta, \varphi) \mapsto (\vartheta, \varphi)$. In order for $\varphi(\theta,\phi) = \varphi(\theta, \phi+\pi)$, the winding number must be even. Together, these show

$$\deg \in \begin{cases} \mathbb{Z} & \text{for } \delta - s \equiv 0 \,(\text{mod } 8) \\ 2\mathbb{Z} & \text{for } \delta - s \equiv 4 \,(\text{mod } 8) \\ 0 & \text{otherwise}. \end{cases} \tag{B8}$$

This gives a topological understanding of the $\mathbb{Z}$'s and $2\mathbb{Z}$'s on the periodic table in terms of winding number, which can be identified with the more general analytic invariants, namely, Chern numbers for nonchiral classes

$$n = \frac{1}{\left(\dfrac{d+D}{2}\right)!} \left(\frac{i}{2\pi}\right)^{(d+D)/2} \int_{T^d \times S^D} \text{Tr}(\mathcal{F}^{(d+D)/2}) \tag{B9}$$

and winding numbers of the chiral flipping operator $q(\mathbf{k}, \mathbf{r})$ for chiral ones [see Eqs. (4.1) and (4.2)].

$$n = \frac{\left(\dfrac{d+D-1}{2}\right)!}{(d+D)!(2\pi i)^{(d+D+1)/2}} \int_{T^d \times S^D} \text{Tr}[(q dq^\dagger)^{d+D}]. \tag{B10}$$

The $\mathbb{Z}_2$'s on the periodic table are not directly characterized by winding degree but rather through dimensional reduction.

Given a Hamiltonian $\mathcal{H}(\mathbf{k}, k_1, k_2, \mathbf{r})$ with $s \equiv \delta \mod 8$, its winding degree mod 2 determines the $\mathbb{Z}_2$ classification of its equatorial offspring $\mathcal{H}_{k_2=0}(\mathbf{k}, k_1, \mathbf{r})$ and $\mathcal{H}_{k_{1,2}=0}(\mathbf{k}, \mathbf{r})$. For example, topological insulators in two and three dimensions are equatorial restrictions of a four-dimensional model $\hat{\mathbf{d}}: S^4 \overset{=}{\rightarrow} S^4$ with unit winding number. Around the north pole, the Hamiltonian has the form

$$\mathcal{H}(\mathbf{k}, k_4) = (m + \varepsilon k^2)\mu_1 + \mathbf{k} \cdot \mu_3 \vec{\sigma} + k_4 \mu_2 \qquad \text{(B11)}$$

and on the equator $k_4 = 0$, this gives a three-dimensional Dirac theory of mass m around $\mathbf{k} = \mathbf{0}$ that locally describes 3D topological insulators Bi_2Se_3 and Bi_2Te_3 around Γ.

APPENDIX C: ZERO MODES IN THE HARMONIC OSCILLATOR MODEL

We present exact solvable soliton states of Dirac-type defect Hamiltonians. These include zero modes at a point defect of a Hamiltonian in the chiral class AIII and chiral modes along a line defect of a Hamiltonian in the nonchiral class A. We establish the connection between the two kinds of boundary modes through the Hamiltonian mapping in Eq. (A1).

A nontrivial chiral Hamiltonian isotropic around a point defect at $\mathbf{r} = 0$ is a Dirac operator

$$\mathcal{H} = -i\vec{\gamma} \cdot \nabla + \mathbf{r} \cdot \vec{\Gamma}, \qquad \text{(C1)}$$

where the chiral operator is $\Pi = i^d \Pi_{j=1}^d \gamma_j \Gamma_j$ and its adiabatic limit $e^{-i\mathbf{k} \cdot \mathbf{r}} \mathcal{H} e^{i\mathbf{k} \cdot \mathbf{r}} = \mathbf{k} \cdot \vec{\gamma} + \mathbf{r} \cdot \vec{\Gamma}$ has unit winding degree on $S^{2d-1} = \{(\mathbf{k}, \mathbf{r}) : k^2 + r^2 = 1\}$.

$$\mathcal{H}^2 = -\nabla^2 + r^2 - i\vec{\gamma} \cdot \vec{\Gamma} \qquad \text{(C2)}$$

and the spectrum is determined by the quantum numbers $n_j \geq 0$ of the harmonic oscillator and the parities ξ_j of the mutually commuting matrices $i\gamma_j \Gamma_j$, for $j = 1, \ldots, d$.

$$\mathcal{E}^2 = \sum_{j=1}^d 2n_j + 1 - \xi_j. \qquad \text{(C3)}$$

The unique zero-energy state $|\Psi_0\rangle$, indexed by $n_j = 0$ and $\xi_j = 1$, has positive chirality $\Pi = +1$, and is exponentially localized at the point defect as $\Psi_0(\mathbf{r}) \propto e^{-1/2r^2}$.

Next we consider a nonchiral Hamiltonian isotropic along a line defect.

$$\mathcal{H}(k_\parallel) = k_\parallel \Pi - i\vec{\gamma} \cdot \nabla + \mathbf{r} \cdot \vec{\Gamma}, \qquad \text{(C4)}$$

where $k_\parallel$ is parallel to the defect line, $\mathbf{r}$ and ∇ are normal position and derivative. Its adiabatic limit $e^{-i\mathbf{k} \cdot \mathbf{r}} \mathcal{H}(k_\parallel) e^{i\mathbf{k} \cdot \mathbf{r}} = k_\parallel \Pi + \mathbf{k} \cdot \vec{\gamma} + \mathbf{r} \cdot \vec{\Gamma}$ is related to that of Eq. (C1) by (1,1) periodicity and has unit winding degree on $S^{2d} = \{(k_\parallel, \mathbf{k}, \mathbf{r}) : k_\parallel^2 + |\mathbf{k}|^2 + |\mathbf{r}|^2 = 1\}$. The zero mode $|\Psi_0\rangle$ of Eq. (C1) gives rise to a positive chiral mode, $\mathcal{H}(k_\parallel)|\Psi_0\rangle = k_\parallel \Pi |\Psi_0\rangle = +k_\parallel |\Psi_0\rangle$ (Fig. 12).

The two examples verified bulk-boundary correspondence through identifying analytic information of the defect-bound solitons and the topology of slowly spatial modulated theo-
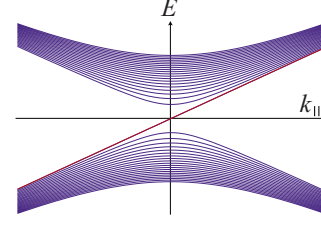

FIG. 12. (Color online) Energy spectrum of Hamiltonian (C4). The zero mode $|\Psi_0\rangle$ of positive chirality at $k_\parallel = 0$ corresponds the chiral mode that generates the midgap $k_\parallel$-linear energy spectrum.

ries far away from the defect. The single zero mode of Eq. (C1) and spectral flow of Eq. (C4) are equated to unit winding degree of an adiabatic limit. In general, bulk-boundary correspondence is mathematically summarized by index theorems that associate certain analytic and topological indices of Hamiltonians.[41,45,46,56,86,87]

APPENDIX D: INVARIANT FOR POINT DEFECTS IN CLASS D AND BDI

We follow the derivation given in Ref. 53, which was based on Qi, Hughes, and Zhang's formulation of the topological invariant characterizing a three-dimensional topological insulator.[10] For a point defect in d dimensions, the Hamiltonian $\mathcal{H}(\mathbf{k}, \mathbf{r})$ depends on d momentum variables and $d-1$ position variables. We introduce a one parameter deformation $\widetilde{\mathcal{H}}(\lambda, \mathbf{k}, \mathbf{r})$ that connects $\widetilde{\mathcal{H}}(\mathbf{k}, \mathbf{r})$ at $\lambda = 0$ to a constant Hamiltonian at $\lambda = 1$ while breaking particle-hole symmetry. The particle-hole symmetry can be restored by including a mirror image $\widetilde{\mathcal{H}}(\lambda, \mathbf{k}, \mathbf{r}) = -\Xi \mathcal{H}(-\lambda, \mathbf{k}, \mathbf{r})\Xi^{-1}$ for $-1 < \lambda < 0$. For $\lambda = \pm$, $(\mathbf{k}, \mathbf{r})$ can be replaced by a single point, so the $2d$ parameter space $(\lambda, \mathbf{k}, \mathbf{r})$ is the suspension $\Sigma(T^d \times S^{d-1})$ of the original space. The Hamiltonian defined on this space is characterized by its dth Chern character

$$\nu = \frac{1}{d!}\left(\frac{i}{2\pi}\right)^d \int_{\Sigma(T^d \times S^{d-1})} \text{Tr}[\mathcal{F}^d]. \qquad \text{(D1)}$$

Due to particle-hole symmetry, the contributions from the two hemispheres $\lambda > 0$, $\lambda < 0$ are equal. Using the fact that the integrand is the derivative of the Chern Simons form, $\text{Tr}[\mathcal{F}^d] = d\mathcal{Q}_{2d-1}$, we can therefore write

$$\nu = \frac{2}{d!}\left(\frac{i}{2\pi}\right)^d \int_{T^d \times S^{d-1}} \mathcal{Q}_{2d-1}. \qquad \text{(D2)}$$

As was the case in Refs. 10 and 53, ν can be different for different deformations $\mathcal{H}(\lambda, \mathbf{k}, \mathbf{r})$. However, particle-hole symmetry requires the difference is an even integer. Thus, the parity of Eq. (D2) defines the $\mathbb{Z}_2$ invariant.

The Chern Simons form $\mathcal{Q}_{2d-1}$ can be expressed in terms of the connection $\mathcal{A}$ via the general formula
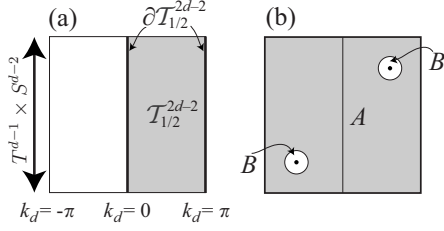
FIG. 13. (a) Schematic of the base space $\mathcal{T}^{2d-2}=T^d\times S^{d-2}$, split into two halves. (b) Division of $\mathcal{T}^{2d-2}$ into patches A and B, each closed under TR and has an individual valence frame $|u_m^{A/B}\rangle$ that satisfies the gauge condition in Eq. (E2).

$$\mathcal{Q}_{2d-1} = d\int_0^1 dt\, \text{Tr}[\mathcal{A}(td\mathcal{A}+t^2\mathcal{A}^2)^{d-1}]. \qquad (D3)$$

In the addition of time-reversal symmetry $\Theta^2=1$ or equivalently a chiral symmetry $\Pi=\Theta\Pi=\tau_z$, a valence frame of the BDI Hamiltonian (4.1) can be chosen to be

$$u(\mathbf{k},\mathbf{r})=\frac{1}{\sqrt{2}}\begin{pmatrix} q(\mathbf{k},\mathbf{r}) \\ -\mathbb{1} \end{pmatrix}, \qquad (D4)$$

where q is unitary and $\mathbb{1}$ is the identity matrix. This corresponds the Berry connection $\mathcal{A}=u^\dagger du=\frac{1}{2}q^\dagger dq$ and Chern-Simons form

$$Q_{2d-1}=\frac{d}{2}\int_0^1 dt\left[\frac{t}{2}\left(\frac{t}{2}-1\right)\right]^{d-1}\text{Tr}[(q^\dagger dq)^{2d-1}]$$

$$=\frac{(-1)^d}{2}\frac{d!(d-1)!}{(2d-1)!}\text{Tr}[(q^\dagger dq)^{2d-1}]. \qquad (D5)$$

This equates the winding number of q in Eq. (4.2) to the Chern Simons invariant in Eq. (4.9).

APPENDIX E: INVARIANT FOR LINE DEFECTS IN CLASS AII AND POINT DEFECTS IN DIII

We formulate a topological invariant that characterizes line defects in class AII in all dimensions that is analogous to the integral formula invariant characterizing the quantum spin-Hall insulator introduced in Ref. 73. This can be applied to weak topological insulators in three dimensions with dislocation around a line defect. The invariant can be indirectly applied to strong topological insulators through decomposition into strong and weak components. As a consequence of the Hamiltonian mapping in Eq. (A1) that identifies (s =4, δ=2) and (s=3, δ=1), this gives a new topological invariant that classified point defects in class DIII in all dimensions.

1. Line defects in class AII

The base space manifold is $\mathcal{T}^{2d-2}=T^d\times S^{d-2}$, where T^d is the Brillouin zone and $S^{d-2}\times\mathbb{R}$ is a cylindrical neighborhood that wraps around the line defect in real space. Divide the base space into two pieces, $\mathcal{T}^{2d-2}_{1/2}$ and its time reversal counterpart [see Fig. 13(a)]. We will show the $\mathbb{Z}_2$ invariant

$$\nu=\frac{i^{d-1}}{(d-1)!(2\pi)^{d-1}}\left[\int_{\mathcal{T}^{2d-2}_{1/2}}\text{Tr}(\mathcal{F}^{d-1})-\oint_{\partial\mathcal{T}^{2d-2}_{1/2}}\mathcal{Q}_{2d-3}\right] \qquad (E1)$$

topologically classifies line defects in AII, where the Chern-Simons form, defined by Eq. (D3), is generated by the Berry connection $\mathcal{A}_{mn}=\langle u_m(\mathbf{k},\mathbf{r})|du_n(\mathbf{k},\mathbf{r})\rangle$, and the valence frame $u_m(\mathbf{k},\mathbf{r})$ satisfies the gauge condition

$$w_{mn}(\mathbf{k},\mathbf{r})=\langle u_m(\mathbf{k},\mathbf{r})|\Theta u_n(-\mathbf{k},\mathbf{r})\rangle=\text{constant} \qquad (E2)$$

on the boundary $(\mathbf{k},\mathbf{r})\in\partial\mathcal{T}^{2d-2}_{1/2}$.

The nontriviality of the $\mathbb{Z}_2$ invariant is a topological obstruction to choosing a global continuous valence frame $|u_m(\mathbf{k},\mathbf{r})\rangle$ that satisfies the gauge condition in Eq. (E2) on the whole base space $\mathcal{T}^{2d-2}$. If there is no topological obstruction from the bulk,[88] the gauge condition forces the valence frame to be singular at two points, depicted in Fig. 13(b), related to each other by time reversal. One removes the singularity by picking another valence frame locally defined on two small balls enclosing the two singular points, denoted by B in Fig. 13(b). We therefore have two valence frames $|u_m^{A/B}(\mathbf{k},\mathbf{r})\rangle$ defined on two patches of the base space, $A=\mathcal{T}^{2d-2}\backslash B$ and B, each obeying the gauge condition in Eq. (E2).

The wave functions on the two patches translate into each other through transition function

$$t_{mn}^{AB}(\mathbf{k},\mathbf{r})=\langle u_m^A(\mathbf{k},\mathbf{r})|u_n^B(\mathbf{k},\mathbf{r})\rangle\in U(k) \qquad (E3)$$

on the boundary $\partial B\approx S^{2d-3}\cup S^{2d-3}$. The function behavior on the two disjoint $(2d-3)$ spheres is related by time reversal. The topology is characterized by the winding of $t^{AB}:S^{2d-3}\to U(k)$ on one of the spheres

$$\nu=\frac{(d-2)!}{(2d-3)!(2\pi i)^{d-1}}\oint_{S^{2d-3}}\text{Tr}[(t^{AB}d(t^{AB})^\dagger)^{2d-3}], \quad (E4)$$

$$=\frac{(-1)^d}{(d-1)!(2\pi i)^{d-1}}\oint_{S^{2d-3}}(\mathcal{Q}_{2d-3}^A-\mathcal{Q}_{2d-3}^B). \qquad (E5)$$

The two integrals can be evaluated separately. Since $d\mathcal{Q}_{2d-3}=\text{Tr}(\mathcal{F}^{d-1})$, Stokes' theorem tells us

$$\int_{A\cap\mathcal{T}^{2d-2}_{1/2}}\text{Tr}(\mathcal{F}^{d-1})=(\oint_{\partial\mathcal{T}^{2d-2}_{1/2}}-\oint_{S^{2d-3}})\mathcal{Q}^A_{2d-3},$$

$$\int_{B\cap\mathcal{T}^{2d-2}_{1/2}}\text{Tr}(\mathcal{F}^{d-1})=\oint_{S^{2d-3}}\mathcal{Q}^B_{2d-3}.$$

Combining these into Eq. (E5) identifies the $\mathbb{Z}_2$ invariant in Eq. (E1) with the winding number of the transition function.

The curvature term in Eq. (E1) is gauge invariant. Any gauge transformation on the boundary $\partial\mathcal{T}^{2d-2}_{1/2}$ respecting the gauge condition in Eq. (E2) has even winding number and would alter the Chern-Simons integral by an even integer. The gauge condition is therefore essential to make the formula nonvacuous.

Spin Chern number

A quantum spin-Hall insulator is characterized by its spin Chern number $n_\sigma = (n_\uparrow - n_\downarrow)/2$. We generalize this to time-reversal invariant line defects of all dimensions by equating it with Eq. (E1). This applies in particular to a model we considered for a linear Josephson junction in Sec. III D.

A spin operator S is a unitary operator, square to unity, commutes with the Hamiltonian, and anticommutes with the time reversal operator. The valence spin frame

$$|u_m^\downarrow(\mathbf{k},\mathbf{r})\rangle = \Theta|u_m^\uparrow(-\mathbf{k},\mathbf{r})\rangle \qquad (E6)$$

automatically satisfies the time reversal gauge constraint in Eq. (E2). It is straightforward to check that the curvature and Chern-Simons form can be split as direct sums according to spins.

$$\mathcal{F}(\mathbf{k},\mathbf{r}) = \mathcal{F}^\uparrow(\mathbf{k},\mathbf{r}) \oplus \mathcal{F}^\uparrow(-\mathbf{k},\mathbf{r})^*, \qquad (E7)$$

$$\mathcal{Q}_{2d-3}(\mathbf{k},\mathbf{r}) = \mathcal{Q}_{2d-3}^\uparrow(\mathbf{k},\mathbf{r}) \oplus \mathcal{Q}_{2d-3}^\uparrow(-\mathbf{k},\mathbf{r})^*. \qquad (E8)$$

Again assuming that there is no lower dimensional "weak" topology, the $\uparrow$ frame can be defined everywhere on $\mathcal{T}^{2d-2}$ with a singularity at one point, say, in $\mathcal{T}_{1/2}^{2d-2}$, and the $\downarrow$ frame is singular only at the time reversal of that point.

The curvature term of Eq. (E1) splits into two terms

$$\int_{\mathcal{T}_{1/2}^{2d-2}} \mathrm{Tr}(\mathcal{F}^{d-1}) = \left[\int_{\mathcal{T}_{1/2}^{2d-2}} - \int_{\mathcal{T}^{2d-2}\backslash\mathcal{T}_{1/2}^{2d-2}}\right] \mathrm{Tr}(\mathcal{F}_\uparrow^{d-1}) \quad (E9)$$

and the two spin components of the Chern-Simons term $\oint_{\partial\mathcal{T}_{1/2}^{2d-2}} \mathcal{Q}_{2d-3}$ add up into

$$2\oint_{\partial\mathcal{T}_{1/2}^{2d-2}} \mathcal{Q}_{2d-3}^\uparrow = -2\int_{\mathcal{T}^{2d-2}\backslash\mathcal{T}_{1/2}^{2d-2}} \mathrm{Tr}(\mathcal{F}_\uparrow^{d-1}) \qquad (E10)$$

by Stokes theorem.

Combining these two, we equate Eq. (E1) to the spin Chern number

$$n_\uparrow = \frac{i^{d-1}}{(d-1)!2\pi^{d-1}} \int_{\mathcal{T}^{2d-2}} \mathrm{Tr}(\mathcal{F}_\uparrow^{d-1}). \qquad (E11)$$

Time reversal requires $n_{tot} = n_\uparrow + n_\downarrow = 0$ and therefore $n_\sigma = (n_\uparrow - n_\downarrow)/2 = n_\uparrow$.

2. Point defects in class DIII

The base space manifold is $\mathcal{T}^{2d-1} = T^d \times S^{d-1}$. The Hamiltonian mapping in Eq. (A1) relates a point-defect Hamiltonian $\mathcal{H}(\mathbf{k},\mathbf{r})$ in class DIII to a line-defect Hamiltonian $\mathcal{H}(\mathbf{k},\mathbf{r},\theta) = \cos\theta\mathcal{H}(\mathbf{k},\mathbf{r}) + \sin\theta\Pi$ in class AII, where $\Pi = i\Theta\Xi$ is the chiral operator, $(\mathbf{k},\mathbf{r},\theta) \in \Sigma\mathcal{T}^{2d-1}$ [see Fig. 14(a)] and θ is odd under time reversal. The line defect Hamiltonian $\mathcal{H}(\mathbf{k},\mathbf{r},\theta)$ is topologically characterized by the generalization of Eq. (E1), which was proven to be identical to the winding number in Eq. (E4) of the transition function t^{AB} [see Fig. 14(b) for the definition of patches A and B]. We will utilize this to construct a topological invariant that characterizes point defects in class DIII.
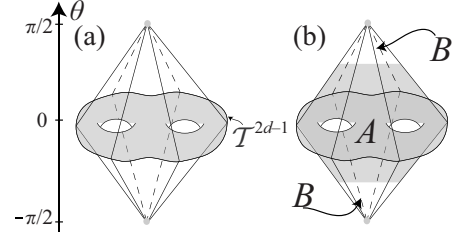

FIG. 14. (a) Schematic of the suspension $\Sigma\mathcal{T}^{2d-1}$. (b) Decomposition into patches A and B.

Set $\Theta = \tau_y K$ and $\Xi = \tau_x K$ under an appropriate choice of basis. A *canonical* valence frame of $\mathcal{H}(\mathbf{k},\mathbf{r},\theta)$ can be chosen to be

$$u_+^B(\mathbf{k},\mathbf{r},\theta) = \begin{pmatrix} \sin\left(\dfrac{\pi}{4} - \dfrac{\theta}{2}\right)q(\mathbf{k},\mathbf{r}) \\[2mm] -\cos\left(\dfrac{\pi}{4} - \dfrac{\theta}{2}\right)\mathbb{1} \end{pmatrix}, \qquad (E12)$$

where $q(\mathbf{k},\mathbf{r}) \in U(k)$ is from the canonical form of the chiral Hamiltonian $\mathcal{H}(\mathbf{k},\mathbf{r})$ in Eq. (4.1), $\mathbb{1}$ is the $k \times k$ identity matrix, and the valence frame is nonsingular everywhere except at $\theta = -\pi/2$. There is a gauge transformation $u_+^B \to u^A = u_+^B t^{BA}$ everywhere except $\theta = \pm\pi/2$ such that the new frame u^A satisfies the gauge condition in Eq. (E2).[89] A valence frame on patch B can be constructed by requiring $u_-^B(\theta) = \Theta u_+^B(-\theta)$ around $\theta = -\pi/2$.

The $\mathbb{Z}_2$ topology is characterized by the evenness or oddness of the winding number of t^{AB} as in Eq. (E4). This can be evaluated by the integral along the equator $\theta = 0$

$$\tilde{\nu} = \frac{(d-1)!}{(2d-1)!(2\pi i)^d} \int_{T^d \times S^{d-1}} \mathrm{Tr}\{[t^{AB}d(t^{AB})^\dagger]^{2d-1}\},$$

$$(E13)$$

where $u^A = u_+^B t^{BA}$ is a solution to the gauge condition in Eq. (E2) or equivalently t^{BA} satisfies

$$q(\mathbf{k},\mathbf{r}) = t^{BA}(-\mathbf{k},\mathbf{r})\sigma_y t^{BA}(\mathbf{k},\mathbf{r})^T, \qquad (E14)$$

where the constant in Eq. (E2) is chosen to be $i\sigma_y$.

The winding number in Eq. (E13) can also be expressed as a Chern-Simons integral.

$$\tilde{\nu} = \frac{i^d}{d!(2\pi)^d} \int_{T^d \times S^{d-1}} (\mathcal{Q}_{2d-1}^B - \mathcal{Q}_{2d-1}^A), \qquad (E15)$$

where $\mathcal{Q}^{A/B}$ are the Chern-Simons form generated by valence frames $u^{A/B}$. Restricted to $\theta = 0$, Eq. (E12) gives $u^B(\mathbf{k},\mathbf{r}) = \frac{1}{\sqrt{2}}(q(\mathbf{k},\mathbf{r}),-\mathbb{1})$. Following Eq. (D5), the first term of Eq. (E15) equals half of the winding number of q, which is guaranteed to be zero by time-reversal and particle-hole symmetries. And therefore point defects in DIII are classified by the Chern-Simons invariant

$$\widetilde{\nu} = \frac{1}{d!}\left(\frac{i}{2\pi}\right)^d \int_{T^d \times S^{d-1}} \mathcal{Q}_{2d-1} \bmod 2, \qquad (E16)$$

where the Chern-Simons form is generated by a valence frame that satisfies the time-reversal gauge constraint in Eq. (E2).

Note that the integrality of the Chern-Simons integral in Eq. (E16) is a result of particle-hole symmetry. Forgetting time reversal symmetry, point defects in class D are classified by the Chern-Simons invariant $\nu = 2\widetilde{\nu}$ in Eq. (D2) with a factor of 2. Time-reversal symmetry requires the zero modes to form Kramers doublets and therefore $\nu = 2\widetilde{\nu}$ must be even. A gauge transformation in general can alter $\widetilde{\nu}$ by any integer. Thus, similar to the formula in class AII, the time-reversal gauge constraint in Eq. (E2) is essential so that Eq. (E16) is nonvacuous.

Fixed points formula in 1D

We here identify Eq. (E16), or equivalently Eq. (E13), to a fixed point invariant in 1 dimension. In Ref. 82, Qi, Hughes, and Zhang showed that 1D time reversal invariant superconductors are $\mathbb{Z}_2$ classified by the topological invariant

$$(-1)^{\widetilde{\nu}} = \frac{\mathrm{Pf}(q_{k=\pi})}{\mathrm{Pf}(q_{k=0})}\exp\left[\frac{1}{2}\int_0^\pi \mathrm{Tr}(q_k dq_k^\dagger)\right] \qquad (E17)$$

under the basis $\Theta = \tau_y K$ and $\Xi = \tau_x K$, where q_k is the chiral flipping operator in Eq. (4.1). Time-reversal and particle-hole symmetries requires $q_k = -q_{-k}^T$. Hence the Pfaffians are well defined as q_k is antisymmetric at the fixed points $k = 0, \pi$.

Using the gauge condition in Eq. (E14), we can expressed the Pfaffians as $\mathrm{Pf}(\Theta q_{k=0,\pi}) = \det(t_{k=0,\pi})\mathrm{Pf}(\sigma_y)$, where t_k^{BA} is abbreviated to t_k.

$$\frac{\mathrm{Pf}(q_{k=\pi})}{\mathrm{Pf}(q_{k=0})} = \exp\left[-\int_0^\pi \mathrm{Tr}(t_k dt_k^\dagger)\right]. \qquad (E18)$$

Substitute Eq. (E14) into the Cartan form $\mathrm{Tr}(q_k dq_k^\dagger)$ gives

$$\mathrm{Tr}(q_k dq_k^\dagger) = \mathrm{Tr}(t_{-k} dt_{-k}^\dagger) + \mathrm{Tr}(t_k dt_k^\dagger). \qquad (E19)$$

Combining these into Eq. (E17)

$$(-1)^{\widetilde{\nu}} = \exp\left[\frac{1}{2}\int_0^\pi \mathrm{Tr}(t_{-k} dt_{-k}^\dagger - t_k dt_k^\dagger)\right], \qquad (E20)$$

$$= \exp\left[-\frac{1}{2}\int_{-\pi}^\pi \mathrm{Tr}(t_k dt_k^\dagger)\right], \qquad (E21)$$
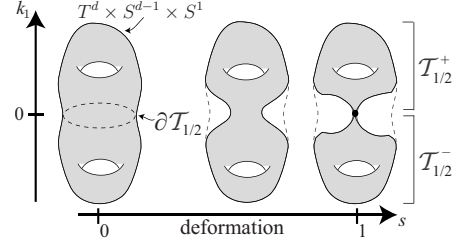
which agrees Eq. (E13).

FIG. 15. Deformation of Hamiltonian and its base manifold so that the boundary $\partial \mathcal{T}_{1/2}$ is shrink to a point.

APPENDIX F: INVARIANT FOR FERMION PARITY PUMPS

In the appendix, we will show that a $\mathbb{Z}_2$ invariant in Eq. (5.2) under a particle-hole gauge constraint topologically classified fermion parity pumps in dimension $\delta = 0$ and class D or BDI. (See Sec. V B for the full statement.) We will show Eq. (5.2) using a construction similar to a reasoning in Moore and Balents.[7] We consider a deformation of the Hamiltonian along with the base manifold so that the boundary $\partial \mathcal{T}_{1/2}$ is deformed into a single point (see Fig. 15). Let $s \in [0,1]$ be the deformation variable, and denote $\mathcal{T}_{1/2}(s)$ and $\partial \mathcal{T}_{1/2}(s)$ be the corresponding deformation slices. Chern invariant

$$n = \frac{1}{d!}\left(\frac{i}{2\pi}\right)^d \int_{\mathcal{T}_{1/2}^+(s=1)} \mathrm{Tr}(\mathcal{F}^d) \qquad (F1)$$

integrally classifies Hamiltonians on the half-manifold $\mathcal{T}_{1/2}^+(s=1)$. Particle-hole symmetry requires opposite Chern invariant on the other half. A different choice of deformation could only change the Chern invariant by an even integer.[90] Hence the Chern invariant modulo 2 defines a $\mathbb{Z}_2$ invariant.

The Chern integral can be further deformed and decomposed into

$$\int_{\mathcal{T}_{1/2}^+(s=0)} \mathrm{Tr}(\mathcal{F}^d) + \int_0^1 ds \int_{\partial \mathcal{T}_{1/2}^+(s)} \mathrm{Tr}(\mathcal{F}^d)$$

$$= \int_{\mathcal{T}_{1/2}^+(s=0)} \mathrm{Tr}(\mathcal{F}^d) - \oint_{\partial \mathcal{T}_{1/2}^+(s=0)} \mathcal{Q}_{2d-1}, \qquad (F2)$$

where Stokes' theorem is used and the negative sign is from a change in orientation of the boundary. This proves Eq. (5.2). The particle-hole gauge constraint is built-in since the $G_{\mathbf{k},\mathbf{r},t}(s)$ deforms into a constant at $s=1$ while respecting particle-hole symmetry in Eq. (5.4) at all s.

---

[1] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs, Phys. Rev. Lett. **49**, 405 (1982).

[2] X. L. Qi and S. C. Zhang, Phys. Today **63**(1), 33 (2010).

[3] J. E. Moore, Nature (London) **464**, 194 (2010).

[4] M. Z. Hasan and C. L. Kane, arXiv:1002.3895v1 (unpublished).

[5] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 226801 (2005).

[6] C. L. Kane and E. J. Mele, Phys. Rev. Lett. **95**, 146802 (2005).

[7] J. E. Moore and L. Balents, Phys. Rev. B **75**, 121306(R) (2007).

[8] R. Roy, Phys. Rev. B **79**, 195322 (2009).

[9] L. Fu, C. L. Kane, and E. J. Mele, Phys. Rev. Lett. **98**, 106803 (2007).

[10] X. L. Qi, T. L. Hughes, and S. C. Zhang, Phys. Rev. B **78**, 195424 (2008).

[11] A. Bernevig, T. Hughes, and S. C. Zhang, Science **314**, 1757 (2006).

[12] L. Fu and C. L. Kane, Phys. Rev. B **76**, 045302 (2007).

[13] H. Zhang, C. X. Liu, X. L. Qi, X. Dai, Z. Fang, and S. C. Zhang, Nat. Phys. **5**, 438 (2009).

[14] M. König, S. Wiedmann, C. Brüne, A. Roth, H. Buhmann, L. W. Molenkamp, X.-L. Qi, and S.-C. Zhang, Science **318**, 766 (2007).

[15] M. König, H. Buhmann, L. W. Molenkamp, T. Hughes, C.-X. Liu, X.-L. Qi, and S.-C. Zhang, J. Phys. Soc. Jpn. **77**, 031007 (2008).

[16] A. Roth, C. Brüne, H. Buhmann, L. W. Molenkamp, J. Maciejko, X. L. Qi, and S. C. Zhang, Science **325**, 294 (2009).

[17] D. Hsieh, D. Qian, L. Wray, Y. Xia, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature (London) **452**, 970 (2008).

[18] D. Hsieh, Y. Xia, L. Wray, D. Qian, A. Pal, J. H. Dil, J. Osterwalder, F. Meier, G. Bihlmayer, C. L. Kane, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Science **323**, 919 (2009).

[19] P. Roushan, J. Seo, C. V. Parker, Y. S. Hor, D. Hsieh, D. Qian, A. Richardella, M. Z. Hasan, R. J. Cava, and A. Yazdani, Nature (London) **460**, 1106 (2009).

[20] Y. Xia, D. Qian, D. Hsieh, L. Wray, A. Pal, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nat. Phys. **5**, 398 (2009).

[21] Y. S. Hor, A. Richardella, P. Roushan, Y. Xia, J. G. Checkelsky, A. Yazdani, M. Z. Hasan, N. P. Ong, and R. J. Cava, Phys. Rev. B **79**, 195208 (2009).

[22] Y. L. Chen, J. G. Analytis, J. H. Chu, Z. K. Liu, S. K. Mo, X. L. Qi, H. J. Zhang, D. H. Lu, X. Dai, Z. Fang, S. C. Zhang, I. R. Fisher, Z. Hussain, and Z. X. Shen, Science **325**, 178 (2009).

[23] D. Hsieh, Y. Xia, D. Qian, L. Wray, J. H. Dil, F. Meier, J. Osterwalder, L. Patthey, J. G. Checkelsky, N. P. Ong, A. V. Fedorov, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Nature (London) **460**, 1101 (2009).

[24] S. R. Park, W. S. Jung, C. Kim, D. J. Song, C. Kim, S. Kimura, K. D. Lee, and N. Hur, Phys. Rev. B **81**, 041405(R) (2010).

[25] Z. Alpichshev, J. G. Analytis, J. H. Chu, I. R. Fisher, Y. L. Chen, Z. X. Shen, A. Fang, and A. Kapitulnik, Phys. Rev. Lett. **104**, 016401 (2010).

[26] D. Hsieh, Y. Xia, D. Qian, L. Wray, F. Meier, J. H. Dil, J. Osterwalder, L. Patthey, A. V. Fedorov, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan, Phys. Rev. Lett. **103**, 146401 (2009).

[27] A. Shitade, H. Katsura, J. Kunes, X. L. Qi, S. C. Zhang, and N. Nagaosa, Phys. Rev. Lett. **102**, 256403 (2009).

[28] R. Li, J. Wang, X. L. Qi, and S. C. Zhang, Nat. Phys. **6**, 376 (2010).

[29] S. Chadov, X. L. Qi, J. Kübler, G. H. Fecher, C. Felser, and S. C. Zhang, Nature Mater. **9**, 541 (2010).

[30] H. Lin, L. A. Wray, Y. Xia, S. Xu, S. Jia, R. J. Cava, A. Bansil, and M. Z. Hasan, Nature Mater. **9**, 546 (2010).

[31] H. Lin, L. A. Wray, Y. Xia, S. Y. Xu, S. Jia, R. J. Cava, A. Bansil, and M. Z. Hasan, arXiv:1003.2615 (unpublished).

[32] H. Lin, L. A. Wray, Y. Xia, S. Y. Xu, S. Jia, R. J. Cava, A. Bansil, and M. Z. Hasan, arXiv:1004.0999 (unpublished).

[33] B. Yan, C. X. Liu, H. J. Zhang, C. Y. Yam, X. L. Qi, T. Frauenheim, and S. C. Zhang, arXiv:1003.0074 (unpublished).

[34] R. Roy, arXiv:0803.2868 (unpublished).

[35] A. P. Schnyder, S. Ryu, A. Furusaki, and A. W. W. Ludwig, Phys. Rev. B **78**, 195125 (2008); AIP Conf. Proc. **1134**, 10 (2009).

[36] A. Kitaev, AIP Conf. Proc. **1134**, 22 (2009).

[37] X. L. Qi, T. L. Hughes, S. Raghu, and S. C. Zhang, Phys. Rev. Lett. **102**, 187001 (2009).

[38] A. Kitaev, arXiv:cond-mat/0010440 (unpublished).

[39] N. Read and D. Green, Phys. Rev. B **61**, 10267 (2000).

[40] A. P. Mackenzie and Y. Maeno, Rev. Mod. Phys. **75**, 657 (2003).

[41] G. E. Volovik, *The Universe in a Helium Droplet* (Clarendon, Oxford, 2003).

[42] G. E. Volovik, JETP Lett. **90**, 587 (2009).

[43] P. Hořava, Phys. Rev. Lett. **95**, 016405 (2005).

[44] A. Altland and M. R. Zirnbauer, Phys. Rev. B **55**, 1142 (1997).

[45] R. Jackiw and C. Rebbi, Phys. Rev. D **13**, 3398 (1976).

[46] R. Jackiw and P. Rossi, Nucl. Phys. B **190**, 681 (1981).

[47] W. P. Su, J. R. Schrieffer, and A. J. Heeger, Phys. Rev. Lett. **42**, 1698 (1979).

[48] L. Fu and C. L. Kane, Phys. Rev. Lett. **100**, 096407 (2008).

[49] J. D. Sau, R. M. Lutchyn, S. Tewari, and S. Das Sarma, Phys. Rev. Lett. **104**, 040502 (2010).

[50] J. Alicea, Phys. Rev. B **81**, 125318 (2010).

[51] R. M. Lutchyn, J. D. Sau, and S. Das Sarma, Phys. Rev. Lett. **105**, 077001 (2010).

[52] Y. Oreg, G. Refael, and F. von Oppen, arXiv:1003.1145 (unpublished).

[53] J. C. Y. Teo and C. L. Kane, Phys. Rev. Lett. **104**, 046401 (2010).

[54] M. Nakahara, *Geometry, Topology and Physics* (Adam Hilger, Bristol, 1990).

[55] J. Goldstone and F. Wilczek, Phys. Rev. Lett. **47**, 986 (1981).

[56] E. J. Weinberg, Phys. Rev. D **24**, 2669 (1981).

[57] E. Witten, Phys. Lett. **117B**, 324 (1982).

[58] E. Witten, J. Diff. Geom. **17**, 611 (1982).

[59] S. C. Davis, A. C. Davis, and W. B. Perkins, Phys. Lett. **408B**, 81 (1997).

[60] D. J. Thouless, Phys. Rev. B **27**, 6083 (1983).

[61] Q. Niu and D. J. Thouless, J. Phys. A **17**, 2453 (1984).

[62] M. Freedman, M. B. Hastings, C. Nayak, X. L. Qi, K. Walker, and Z. Wang, arXiv:1005.0583 (unpublished).

[63] M. Karoubi, *K-Theory: An Introduction* (Springer-Verlag, Berlin, 1978).

[64] H. B. Lawson and M. L. Michelsohn, *Spin Geometry* (Princeton University Press, Princeton, 1989).

[65] M. Atiyah, *K Theory* (Westview Press, Boulder, CO, 1994).

[66] R. Bott, Ann. Math. **70**, 313 (1959).

[67] J. Milnor, *Morse Theory* (Princeton University Press, Princeton, 1963).

[68] S. Ryu, A. Schnyder, A. Furusaki, and A. W. W. Ludwig, New J. Phys. **12**, 065010 (2010).

[69] E. I. Blount, Solid State Phys. **13**, 305 (1962).

[70] R. Li, J. Wang, X. L. Qi, and S. C. Zhang, Nat. Phys. **6**, 284 (2010).

[71] R. S. K. Mong, A. M. Essin, and J. E. Moore, Phys. Rev. B **81**, 245209 (2010).

[72] G. E. Volovik, Pis'ma Zh. Eksp. Teor. Fiz. **91**, 61 (2010); arXiv:1001.1514 (unpublished).

[73] L. Fu and C. L. Kane, Phys. Rev. B **74**, 195312 (2006).

[74] T. Fukui and T. Fujiwara, J. Phys. A **42**, 362003 (2009).

[75] Y. Ran, Y. Zhang, and A. Vishwanath, Nat. Phys. **5**, 298 (2009).

[76] P. G. Grinevich and G. E. Volovik, J. Low Temp. Phys. **72**, 371 (1988).

[77] M. A. Silaev and G. E. Volovik, arXiv:1005.4672 (unpublished).

[78] R. B. Laughlin, Phys. Rev. Lett. **80**, 5188 (1998).

[79] L. S. Pontrjagin, Rec. Math. [Mat. Sbornik] N.S. **9**, 331 (1941) [http://mathnet.ru/mst6073].

[80] J. Jäykkä and J. Hietarinta, Phys. Rev. D **79**, 125027 (2009).

[81] L. Kapitanski, London Mathematical Society Durham Symposium (unpublished).

[82] X. L. Qi, T. L. Hughes, and S. C. Zhang, Phys. Rev. B **81**, 134508 (2010).

[83] L. Fu, and C. L. Kane, Phys. Rev. B **79**, 161408(R)(2009).

[84] Y. Ran, P. Hosur, and A. Vishwanath, arXiv:1003.1964 (unpublished).

[85] R. B. Laughlin, Phys. Rev. B **23**, 5632(R) (1981).

[86] T. Fukui and T. Fujiwara, J. Phys. Soc. Jpn. **79**, 033701 (2010).

[87] T. Fukui, Phys. Rev. B **81**, 214516 (2010).

[88] Bulk topology is a lower dimenional topological obstruction coming purely from the momentum part T^d of the base space $T^d \times S^{d-2}$. One could mathematically remove the bulk topological obstruction by adding a defectless Hamiltonian $\widetilde{\mathcal{H}}(\mathbf{k},\mathbf{r}) = \mathcal{H}(\mathbf{k},\mathbf{r}) \oplus \mathcal{H}_0(\mathbf{k})$.

[89] Again, assume the bulk has trivial topology. Or otherwise remove the topological obstruction mathematically by adding a defectless Hamiltonian.

[90] Deformation of Hamiltonians on $\partial \mathcal{T}_{1/2}(s)$, for $s \in [0,1]$ has dimension $\delta = (d-1) - (d-1+1+1) = -2$, and is classified by $2\mathbb{Z}$.