PHYSICS OF THE AXION COUPLING

by

NICODEMOS VARNAVA

A dissertation submitted to the
School of Graduate Studies
Rutgers, The State University of New Jersey
In partial fulfillment of the requirements
For the degree of
Doctor of Philosophy
Graduate Program in Physics and Astronomy

Written under the direction of
David Vanderbilt
and approved by

__________________________
__________________________
__________________________
__________________________
__________________________

New Brunswick, New Jersey
October, 2021
ABSTRACT OF THE DISSERTATION

Physics of the axion coupling

By Nicodemos Varnava
Dissertation Director:
David Vanderbilt

The application of topology in condensed matter physics has transformed the field by introducing new phases of matter. The prime example is the Chern number, an integer invariant of 2D manifolds, that is associated with the ground state of 2D insulators and distinguishes them based on their quantized Hall response. Physically, this quantization can be understood by viewing the Hall response as the change of the 1D polarization after a slow cyclic evolution. The result of which is necessarily the transport of an integer number electrons over one lattice constant. Importantly, Chern invariants are defined for every even dimensional manifold. In particular, the second Chern number manifests as a quantized nonlinear response of 4D insulators. In this case, the analog of the 1D polarization plays the 3D Chern-Simons axion coupling which describes an isotropic magnetoelectric effect. When time-reversal or inversion symmetry are present the response becomes quantized defining strong topological insulators and axion insulators respectively.

In this thesis, we explore the physics of the axion coupling with an emphasis on magnetic topological insulators. We make contributions to the field by elucidating their physical properties, uncovering their mathematical manifestation and predicting material candidates using density function theory.
Specifically, we construct a computational tool to formally answer how the integer part of the surface anomalous Hall conductivity of an axion insulator is determined. We then uncover phenomena at the surfaces of axion insulators including a novel quantum point junction. Next we consider how the axion coupling manifests in the hybrid Wannier representation. After describing the general features, we focus on quantized axion coupling in the presence of symmetries. We classify all such symmetries and explain how each class manifests. Following that we take a closer look at a specific application of antiferromagnetic topological insulators, the quantum point junction. We show how it realizes a robustly controllable quantum gate with applications in electron quantum optics. Finally, we discuss our search for an axion insulator, and present our preliminary results which suggest that the Zintl compounds Eu$_5$Ga$_2$Sb$_6$ and Eu$_5$Tl$_2$Sb$_6$ are promising candidates.
Acknowledgements

First, I want to thank my advisor David Vanderbilt, for teaching me how to think about physics. More importantly, because he gave me a model of a person to strive to be. Except from being a genius, David is virtuous, caring and leads by example. I was very lucky in having him as my PhD mentor.

I would also like to thank my collaborators whom I have great time interacting with and learned so much from, Jed Pixley, Justin Wilson, Ivo Souza and Tyrel McQueen. I have also learnt a lot from the amazing community of professors at Rutgers’ Physics and Astronomy department, and I would like to thank Karin Rabe, Donald Hamann, Emil Yuzbashyan, Piers Coleman, David Shih, Greg Moore and Matthew Buckley. I am also grateful to our supporting graduate program director Ronald Gilman and Shirley Hinds.

During these past 5 years at graduate school, I was lucky to have met many amazing people. Among many others I would like to thank Kosta Morfesis, Harry Sims, Victor Drouin-Touchette, Jay Cushing, Aidan Zabalo, Daniel Cerkoney, Ahsan Khan, Konrad Genser, John Bonini, Daniel Seleznev, Shiang Ren, Sobhit Singh and Jinwoong Kim.

A special thanks to the friends I made outside of graduate school. Theo Berbos, Luka Turkalj, Hristiyan Kourtev and Iva Salamon. Iva you are de best.

Finally, I want to thank Maria and George Varnava who I am blessed to call my parents. This work is dedicated to them. Their support gave me the privilege to set and pursue my own goals and their continuing unconditional love is one of the greatest gifts I have in my life.
# Table of Contents

Abstract ................................................................. ii

Acknowledgements ...................................................... iv

1. Introduction .......................................................... 1

2. Background ........................................................... 7
   2.1. Preliminaries ..................................................... 7
       2.1.1. Berry phase .................................................. 7
       2.1.2. Wannier representation .................................... 11
       2.1.3. Hybrid Wannier representation ............................ 12
   2.2. Berry phase in electronic structure .......................... 15
       2.2.1. Electric polarization ........................................ 15
       2.2.2. Quantum anomalous Hall effect ........................... 17
       2.2.3. Adiabatic pumping ......................................... 20
   2.3. Axion coupling in electronic structure ....................... 22
       2.3.1. From the second Chern number to the axion coupling ... 23
       2.3.2. Axion electrodynamics ..................................... 27
       2.3.3. Topological insulators ..................................... 28

3. Surfaces of axion insulators ....................................... 34
   3.1. Calculation of the surface anomalous Hall conductivity .... 35
   3.2. Minimal model on the pyrochlore lattice ..................... 37
       3.2.1. The pyrochlore lattice: A playground for topological phenomena 38
       3.2.2. The model .................................................. 38
   3.3. Results ......................................................... 40
3.3.1. All-in-all-out spin configuration ............................................. 41
3.3.2. Ferromagnetic spin configuration ............................................. 47
3.4. Surface Phenomena .............................................................. 48
3.4.1. Surface AHC and magnetization direction ................................. 48
3.4.2. Termination-dependent surface AHC in the AIAO configuration .... 50
3.4.3. Higher-order topological classification of an FM axion insulators 51

4. Axion coupling in the hybrid Wannier representation ..................... 55
  4.1. General properties .............................................................. 57
    4.1.1. Disconnected Wannier band structure .................................. 58
    4.1.2. Introduction of a cutting surface ....................................... 62
    4.1.3. Connected Wannier band structure ..................................... 64
  4.2. Symmetry considerations .................................................... 67
    4.2.1. Axion-odd insulators ..................................................... 67
    4.2.2. Determination of the quantized axion coupling ....................... 70
  4.3. Bulk-boundary correspondence .............................................. 81
    4.3.1. Protected flow of the surface energy bands ............................ 82
    4.3.2. Protected flow of the bulk Wannier bands ............................. 84
    4.3.3. Higher-order topology ................................................... 85
  4.4. Special cases and numerical tests ......................................... 86
    4.4.1. Time-reversal symmetry ................................................ 87
    4.4.2. Inversion symmetry .................................................... 90
    4.4.3. Mirror symmetry ....................................................... 94
    4.4.4. Glide mirror symmetry ................................................ 98

5. Controllable quantum point junction on the surface of an antiferro-
magnetic topological insulator .................................................. 102
  5.1. Introduction ........................................................................ 102
  5.2. Results .............................................................................. 105
    5.2.1. Extracting the $S$ matrix .................................................. 105
5.2.2. Controlling the $S$ matrix ........................................ 110
5.2.3. Robustness of junction ........................................... 112
5.3. Discussion ................................................................. 117
5.4. Methods ................................................................. 119

6. Searching for axion insulators ........................................... 125
   6.1. Introduction ............................................................. 125
   6.2. General approach ..................................................... 127
   6.3. Magnetic topological insulators in Zintl compounds .......... 128
      6.3.1. Prediction of magnetic topological insulator in $\text{Eu}_5\text{(Ga,Tl)}_2\text{Sb}_6$ . 130
      6.3.2. Determination of the magnetic configuration ............... 133
   6.4. Conclusions ............................................................. 137

7. Summary and outlook ...................................................... 138
Chapter 1
Introduction

When Richard Feynman was asked to think of a single sentence that would convey the most important scientific knowledge we possess, he answered simply: “Everything is made of atoms.” This statement is today taken for granted but the foundations of modern science were built on it, from atomic physics and chemistry to biology and modern technologies. Indeed, even if today we know atoms are not the most fundamental unit of reality,\(^1\) we can use them as the starting point to explain the vast majority of observed phenomena.

Condensed matter physics builds upon the concept of atoms by considering large aggregates of them interacting through electromagnetic forces. Founded upon the pillars of statistical and quantum mechanics, it deals with the collective emergent properties of electrons and nuclei. The most generic condensed matter Hamiltonian is derived from the Dirac equation\(^2\) in the non-relativistic limit \((v/c \ll 1)\) and includes the kinetic energy of electrons and nuclei, the electron-electron and electron-nuclei Coulomb interactions, as well as purely relativistic corrections such as the spin-orbit coupling and the Darwin term. In simplifying further the Hamiltonian the nuclear and electronics degrees of freedom can be dealt independently because the characteristic time scale of nuclear dynamics is much slower than that of electrons. In this so called Born–Oppenheimer approximation, one freezes the nuclear degrees of freedom and considers the resulting static lattice and then treats the electron-nuclei Coulomb interactions as an external periodic potential. In this work we will make one further assumption which simplifies

---

\(^1\)Atoms are an emergent concept derived from the Core Theory (a term coined by Frank Wilczek to replace the yawn-inducing name “Standard model”). At the most basic level the Core Theory describes reality through a universal wavefunction of interacting quantum fields.

\(^2\)Usually electromagnetic fields are considered as external degrees of freedom and are not treated in a unified framework as in quantum electron dynamics.
Things greatly but should be applied to materials that are not-too-strongly correlated. Namely, we will assume that electron-electron interactions can be treated on a mean-field level, i.e., the interacting system can be mapped to a system with non-interacting electrons moving in an effective static field generated self-consistently by the other electrons.

We will also restrict our considerations to crystalline solids. These systems are comprised of unit cells that periodically repeat themselves to create macroscopic structures. One reason that condensed matter physicists are very interested in them is that crystals are ubiquitous. Indeed, usually at low temperatures matter spontaneously breaks the rotational symmetry of electromagnetic interactions by choosing one of the minimum energy configurations, effectively lowering the symmetry to a discrete subgroup. More importantly, crystalline solids are fascinating because they enable the prediction of macroscopic properties, that are inherently quantum mechanical.\(^3\) This is because the complexity of the system can depend on the complexity of the unit cell. For (effectively) non-interacting systems this is exemplified by the Bloch theorem which states that solutions \(\psi(r)\) to the single-particle Schrödinger equation in a crystalline potential take the form of a plane wave modulated by a periodic function \(u(r)\),

\[
\psi_{nk}(r) = e^{ik \cdot r} u_{nk}
\]

where \(n\) is an integer associated with the \(N\) degrees of freedom in the unit cell and \(k\) is the crystal momentum associated with the \(N_k\) unit cells. Since the system is not interacting, states at different \(k\) are independent. Remarkably, solving the problem of an \(N \times N_k\)-dimensional electron system has been reduced to solving \(N_k\), \(N\)-dimensional systems, making it computationally tractable. Finally, the collective many-electron wavefunction is constructed by occupying the single particle states \(\psi_{nk}(r)\) and requiring that they obey the Pauli exclusion principle. This is known as the band theory of solids.

In the last century band theory has been a triumph of physics, with the classification of solids into metals, semiconductors and insulators through the concept of an energy gap. In addition, the theory was able to determine the related physical properties in

\footnote{Note that the available states for quantum systems grow exponentially with the number of particles.}
real materials such as electrical resistivity and optical absorption to form the foundation of the understanding of all solid-state devices such as transistors and solar cells, which are the foundation for much of the modern technology.

This was achieved in no small part due to the development of first-principle techniques\(^4\) that enable the approximation of the many-body electron Hamiltonian of real materials with an effective one-particle mean-field Hamiltonian. The prime example of a first-principle technique is the density-functional theory (DFT). DFT is build on the work by Hohenberg and Kohn [50] who proved that the many-body ground-state wave function \(|\Psi\rangle\) and the ground-state energy \(E\) can be regarded as functionals of the ground state electron density \(n(r)\). This then means that the many-body problem can be solved by minimizing this energy functional without reference to the many-body wave function. Subsequently, Kohn and Sham [66] provided an iterative procedure to find such a solution, which importantly depends on the choice of the energy functional. In this way, DFT does not solve the many-body problem but maps it to the problem of choosing the best approximation for the energy functional.\(^5\) Remarkably, such approximations can be made and have resulted in accurate predictions making DFT one the most important tools in simulating real materials.

Even though, the electronic structure theory of crystalline systems is founded on quantum mechanics with the wave function assigning complex amplitudes to different states, almost all quantities of interest can be expressed as Bloch-state expectation values.\(^6\) When computing these quantities, the overall phase in front of any given Bloch function plays no role. However, in the last 30 years, it has become evident that the phases of the electronic states are essential in describing some properties of crystalline materials. These include the electric polarization, the anomalous Hall conductivity and the orbital magnetization. Indeed, even if the quantum mechanical phases cannot

\(^4\)These are simulations that only use the fundamental constants of physics as input to predict physical properties of crystals and molecules.

\(^5\)In particular, it is the exchange-correlation potential, which is part of the energy functional and encodes non-classical many-body effects, that needs to be approximated.

\(^6\)Examples include the band energies, charge densities, spin densities, local densities of states, total energies, forces, stresses, spin magnetic moments, and many more.
be measured in any experiment, the relative phases between nearby states, in some
parametric space such as time $t$, position $r$, the crystal momentum $k$, can have ob-

servable consequences. In particular, the phase acquired by a quantum state when it
is transported along a path contains information of about the geometrical properties
of the parameter space of the Hamiltonian and is known as the geometrical phase or
Berry phase. More generally, this is captured by the theory of fiber bundles, where we
can imagine the parameter space forming the base manifold and at each point on the
manifold we have fibers – in this case in the shape of a circle – representing all available
quantum mechanical phases. This then naturally brings the mathematics of topology
in the discussion since fiber bundles have the property that in any small neighbourhood
of the manifold the space looks like a product space between the manifold and the fiber
but globally the fiber bundles might be distinct structures. For example, a cylinder and
the Möbius strip can both be thought as having a circle for the base manifold and for
a line segment the fiber and even though locally they look the same, globally they are
distinct topological objects.

Similar considerations have more recently let physicists to realize that distinct phases
of matter can be associated with the distinct topological structures of the parameter
space of the Hamiltonians. In addition, by requiring that the Hamiltonian respects
certain symmetries – either onsite or crystalline –, the parameter space takes the shape
of certain manifolds that can then be classified to obtained all the different topological
structures, such as in the example of the cylinder and the Möbius strip. This constituted
a paradigm shift in condensed matter physics as the distinction between different phases
of matter had almost always rested on the Landau-theory of symmetry breaking. In
contrast, two topologically distinct phases have exactly the same symmetries but belong
to classes that differ in their global structure.

An important class of materials that will concern us in this work is that of 3D
insulators. Around the end of the first decade of the 21st century it was theoretically
proposed that crystalline insulators with time-reversal symmetry could be classified by
the so called strong $\mathbb{Z}_2$ index [40, 80], i.e. as either trivial insulators or strong topological
insulators (TI). Strong TIs have the fascinating property that all of their surfaces are
required to have an odd-number of linearly dispersing gapless modes, usually referred to as Dirac cones. Remarkably, soon after their topological classification, using DFT calculations Zhang et al. [152] predicted the crystal Bi$_2$Se$_3$ to be an example of strong TIs and Xia et al. [143] identified it as such by directly observing the Dirac cones on its surface using ARPES measurements. Around the same time Qi et al. [93] discovered that the strong $\mathbb{Z}_2$ index could be determined from a $k$-space integral with the integrand determined by the ground state wave function, the axion magnetoelectric coupling $\theta_{cs}$, with the values $\theta_{cs} = 0, \pi$ associated with the trivial and topological phases respectively. Essin et al. [34] then realized that crystalline symmetries such as inversion or mirror symmetry could also quantize the axion coupling to be 0 or $\pi$. Importantly, a surface of an inversion symmetric insulator, with broken time-reversal is not required to have Dirac cones. Instead if it is gapped, it is required to exhibit a half-integer surface anomalous Hall conductivity, with the integer part being surface dependent. The prediction and experimental identification of such a topological phase, commonly refer to as axion insulators, has turned out to be a much harder task than that of strong TIs.

In this work we explore the physics of the axion coupling with an emphasis on magnetic TIs. These are crystals with symmetries that quantize the axion coupling other than simple time-reversal and have $\theta_{cs} = \pi$. We make significant contributions to the field by elucidating their physical properties and mathematical manifestation. Furthermore, we propose a novel device that harnesses their unique potential and we predict material realizations based on DFT calculations.

The thesis is organized as follows. In Chapter 2, we review the basic mathematical and physical concepts necessary to understand the axion coupling and magnetic TIs. In Chapter 3, we consider axion insulators and construct a computational tool that helps us answer how the integer part of the surface anomalous Hall conductivity is determined. This then enables to uncover novel phenomena at the surfaces of axion insulators. In Chapter 4, we consider how the axion coupling manifests in the hybrid Wannier representation. After describing the general features we classify all axion-odd symmetries and we explain how each class manifests. In Chapter 5, we take a closer look
at a novel junction proposed in Chapter 3 and show how it realizes a robustly controllable quantum gate with applications in electron quantum optics. Finally, in Chapter 6, we discuss our search of an axion insulator using DFT, and present our preliminary results that suggest the Zintl compounds Eu₅(Ga,Tl)₂Sb₆ as potential candidates.
Chapter 2
Background

2.1 Preliminaries

In this section, we introduce some basic mathematical concepts to understand topological phases. First we introduce the concept of Berry’s geometric phase that leads to the notion of a local Berry curvature and the Chern theorem. We then discuss the Wannier representation as an alternative description to the Bloch states. Finally we introduce the hybrid Wannier representation which, as the name suggests, is a hybrid of the Wannier and the Bloch representations.

2.1.1 Berry phase

The “Berry phase” $\phi$ refers to a phase angle that describes the global phase evolution of a complex vector as it is carried around a path in its vector space. In quantum mechanics, the Berry phase arises when one considers cyclic adiabatic evaluations.

Let us consider a physical system described by a Hamiltonian $H(M)$ that depends on time through a set of parameters $q(t) = (q_1(t), q_2(t), \ldots)$ that collectively define a manifold $q \in M$. For example this could be the nuclear coordinates of a moving molecule $r(t) \in M = \mathbb{R}^3$ or more abstractly the crystal momentum of a Bloch state in the presence of a uniform electric field $k(t) = k(0) + e^2/\hbar E t$ with $k(t) \in M = \mathbb{T}^3$.

We consider an adiabatic evolution of the system as $q(t)$ slowly traces a path $C$ in the parameter space. For this this illustrated in Fig. 2.1(a) for a 2D parameter space. The instantaneous energy eigenstates of $H(q)$ at some fixed value of $q$ will then be

$$H(q)|n(q)\rangle = E_n(q)|n(q)\rangle. \quad (2.1)$$

According to the adiabatic theorem if the system is initiated in some state $|n(q(0))\rangle$ it
Figure 2.1: (a) Parameter space $\mathbf{q} = (q_1, q_2)$ and loop $\mathcal{C}$ describing a cyclic adiabatic evolution. (b) Two-dimensional closed manifold with the closed loop $\mathcal{C}$ that is used to show the Chern theorem. Adapted from Vanderbilt [124].

will remain an instantaneous eigenstate of the Hamiltonian throughout the adiabatic evolution. Therefore the only degree of freedom we have is the phase of the quantum state. We write the state at time $t$ using the ansatz

$$|\psi_n(t)\rangle = c_n(t) \exp \left[ - \frac{i}{\hbar} \int_0^t dt' E_n(\mathbf{q}(t')) \right] |n(\mathbf{q}(t))\rangle. \quad (2.2)$$

The second exponential is known as the dynamical phase factor. The factor $c(t)$ allows for the possibility that there may be some extra phase evolution beyond the dynamical phase factor. Plugging the ansatz into the time-dependent Schrödinger equation and multiplying from the left with $\langle n(\mathbf{q}) |$ we find that $c_n(t)$ is a phase factor with

$$c_n(t) = e^{i\phi_n},$$

$$\phi_n = \oint_{\mathcal{C}} d\mathbf{q} \cdot \mathbf{A}_n(\mathbf{q}) \quad (2.3)$$

$$\mathbf{A}_n(\mathbf{q}) = i \langle n(\mathbf{q}) | \frac{\partial}{\partial \mathbf{q}} | n(\mathbf{q}) \rangle.$$  

Here $\phi_n$ is the Berry phase while the vector $\mathbf{A}_n(\mathbf{q})$ is called the Berry connection and contributes an additional phase to the time-evolution. Remarkably, the Berry phase entering into the time-evolving wave function is only a function of the path it has traced in parameter space, and is independent of the rate at which the path is traversed, so long as the parametric evolution is sufficiently slow. Is $\mathbf{A}_n(\mathbf{q})$ then gauge dependent? The answer is yes, if we consider a gauge transformation $|n(\mathbf{q})\rangle \rightarrow e^{i\zeta(\mathbf{q})} |n(\mathbf{q})\rangle$ with $\zeta(\mathbf{q})$ an arbitrary smooth function, $\mathbf{A}_n(\mathbf{q})$ transforms to

$$\mathbf{A}_n(\mathbf{q}) \rightarrow \mathbf{A}_n(\mathbf{q}) - \frac{\partial \zeta(\mathbf{q})}{\partial \mathbf{q}}. \quad (2.4)$$
Consequently, the Berry phase is changed by \( \Delta \phi_n = \zeta(q_i) - \zeta(q_f) \) where \( q_i = q_f \) are the same parameters for a cyclic adiabatic evolution such as the one in Fig. 2.1(a). Does that mean \( \Delta \phi_n = 0 \)? Not necessarily. What we have assumed was that \( |n(q)| \) are smooth with respect to \( q \) and single-valued. This means \( e^{i\zeta(q_i)} = e^{i\zeta(q_f)} \) or equivalently

\[
\zeta(q_i) - \zeta(q_f) = 2\pi n, \quad n \in \mathbb{Z}. \tag{2.5}
\]

This shows that this additional Berry phase can only be changed by an integer multiple of \( 2\pi \) under a gauge transformation and it cannot be removed. This makes the Berry phase or geometric phase a gauge invariant quantity modulo \( 2\pi \).

The loop integral shown in Eq. (2.3) can also be expressed as an area integral of a local gauge-invariant quantity known as the Berry curvature. By applying Stoke’s theorem to the loop integral in Eq. (2.3), one obtains for arbitrary parameter dimensions the Berry phase in terms of the gauge invariant Berry curvature

\[
\phi_n = \oint_C dq_i A^i_n(q) = \frac{1}{2} \int_S dq_i dq_j \Omega^n_{ij} \tag{2.6}
\]

where \( S \) is a subregion in the multi-dimensional parameter space \( q \) enclosed by loop \( C \) and we have switched to the more convenient Einstein convention. In three dimensions Eq. (2.6) can be recast into vector form

\[
\phi_n = \int_S d^2q \cdot \Omega^n, \tag{2.7}
\]

\[
\Omega^n = \nabla_q \times A_n(q),
\]

where \( d^2q = \hat{n} d^2q \) and \( \hat{n} \) is a unit vector normal to the surface element of area \( d^2q \). Here the Berry curvature has become a pseudovector whose \( i \)th component is defined, using the Levi-Civita symbol, as \( \Omega_i = \epsilon^{ijk} \Omega_{jk} \). In the 2D case only \( \Omega_{xy} \) survives and thus becomes a pseudoscalar.

To gain more insight into the physical meaning behind the Berry curvature, we can express it in a form that resembles a response function\(^1\) obtained from perturbation

\(^1\)We will later see this expression emerges in the context of the 2D quantum anomalous Hall effect through the Kubo formula.
theory
\[ \mathbf{\Omega}_{ij}^n = i \sum_{m \neq n} \frac{\langle n| \frac{\partial H}{\partial q_i} | m \rangle \langle m| \frac{\partial H}{\partial q_j} | n \rangle - \langle n| \frac{\partial H}{\partial q_j} | m \rangle \langle m| \frac{\partial H}{\partial q_i} | n \rangle}{(E_n - E_m)^2} \] (2.8)

where to obtain Eq. (2.8), we have used that
\[ \langle m| \frac{\partial}{\partial q_i} | n \rangle = \langle m| \frac{\partial H}{\partial q_i} | n \rangle = \frac{(E_n - E_m)}{E_n - E_m}, \quad \text{if} \quad m \neq n. \] (2.9)

Eq. (2.8) shows that the Berry curvature can be understood as the result of virtual excitations driven by the time evolution of the Hamiltonian. Since in the adiabatic approximation, the system in a certain eigenstate must stay in the same instantaneous eigenstate no first order transition are allowed, i.e. from state \( m \) to state \( n \). However, there could be a virtual process that the system tunnels to the state \( n \) then hops back to the initial state \( m \). The effect of such process is manifested as the Berry curvature, and enters into the phase of the wavefunction after a cyclic evolution.

**Chern theorem** — We end this section by briefly discussing an important theorem in differential topology. It states that the integral of the Berry curvature on any 2D closed manifold is quantized to \( 2\pi \) times an integer \( C \), i.e.,
\[ C = \frac{1}{2\pi} \oint_S d^2 \mathbf{q} \cdot \mathbf{\Omega}. \] (2.10)

This is the famous Chern theorem. The integer \( C \) is known as the Chern number or Chern index of the surface, and can be regarded as a "topological invariant" attached to the manifold of states \( |n(q)\rangle \) defined over the surface \( S \). To see this consider an arbitrary closed line \( C \) on \( S \) that divides the manifold into two regions \( A, B \). This is illustrated in Fig. 2.1(b) for a spherical manifold but the argument is general. Next we ask what is the Berry phase \( \phi \) associated with the boundary between the two regions \( C \) and use Stroke’s theorem we can calculate it using either the Berry curvature over the surface region \( A \) or \( B \). That is we have
\[ \phi = \int_C d\mathbf{q} \cdot \mathbf{A} = \int_A d^2 \mathbf{q} \cdot \mathbf{\Omega} = -\int_B d^2 \mathbf{q} \cdot \mathbf{\Omega}. \] (2.11)

However, as we discussed earlier the two rightmost integrals will be equal only up to a multiple integer of \( 2\pi \). Therefore we have
\[ \int_A d^2 \mathbf{q} \cdot \mathbf{\Omega} + \int_B d^2 \mathbf{q} \cdot \mathbf{\Omega} = 2\pi C \] (2.12)
Figure 2.2: Examples of Wannier functions. (a) Sketch of three adjacent Wannier functions $w_R(x)$ in a 1D crystal of lattice constant $a$. Adapted from [124]. (b) Maximally localized Wannier functions constructed from the s band (left) or from the three p bands (right) of GaAs. Adapted from [78].

which is the Chern theorem.

Remarkably, the Chern invariant can be defined in any even dimension. Of special interest to our forthcoming discussions will be the second Chern number $C^{(2)}$ which classifies a four-dimensional closed manifold $\mathcal{S}$. In terms of the Berry curvature it is given by

$$C^{(2)} = \frac{1}{32\pi^2} \int_{\mathcal{S}} dq \epsilon_{ijkl} \text{Tr}[\Omega_{ij}\Omega_{jk}]. \quad (2.13)$$

2.1.2 Wannier representation

Wannier states offer an alternative description of the ground state of a periodic crystal. Namely, one carries out a unitary transformation from the Bloch states $|\psi_{nk}\rangle = e^{ik\cdot r}|u_{nk}\rangle$ to a set of localized Wannier states

$$|w_{nR}\rangle = \frac{1}{(2\pi)^3} \int_{BZ} dke^{-ik\cdot R} |\psi_{nk}\rangle, \quad (2.14)$$

labeled by a cell index $R$, and a band-like index $n$. The expectation here is that as long as a smooth and periodic gauge can be chosen for the Bloch states $|\psi_{nk}\rangle$, the Wannier functions (WFs) $w_{nR}(r) = \langle r |w_{nR}\rangle$ will be localized at $R$ and decay rapidly away from it. Moreover, since the Fourier transform expressed by Eq. (2.14) is really just a special case of a unitary transformation, the Bloch and Wannier states can be viewed as two different basis sets to describe the same manifold of states associated with some ground
state described by a basis-independent projection operator

\[ P = \sum_{n \in \text{occ}} P_n \]  

(2.15)

where \( P_n \) projects on the band \( n \) and is also basis-independent for an isolated band

\[ P_n = \frac{1}{(2\pi)^3} \int_{BZ} dk |\psi_{nk}| = \sum_R |w_{nR}| \].  

(2.16)

Of course in most crystals, the occupied states become degenerate at high-symmetry points, and as a result the Bloch functions often have a singularity as a function of \( k \) in the vicinity of the degeneracy. To proceed, one can consider a group of bands that are glued together by degeneracies as comprising a composite group and treat this group as a whole. In this case, we need to abandon the notion that each Wannier state should be associated with one and only one energy band. Instead we ask that the set of Wannier states \( \{|w_{nR}\} \) span the same occupied subspace as the Bloch states. Since \( P \) is invariant by an arbitrary \( k \)-dependent unitary transformation to the occupied Bloch states at each \( k \),

\[ |\tilde{\psi}_{nk}\rangle = \sum_m U_{mn}(k) |\psi_{mk}\rangle, \quad n, m \in \text{occ.} \]  

(2.17)

we can use this freedom to smooth out any sharp structure in \( k \)-space that translates into a loss of localization in \( r \)-space before applying Eq. (2.14).

In particular, a general approach to construct localized Wannier functions is to choose \( U_{mn}(k) \) as to minimize the sum of quadratic spreads of the Wannier functions in one unit cell [77] resulting in a set of maximally localized Wannier functions (MLWFs). We will see later it is not always possible to construct MLWFs. Indeed, there can be topological obstruction in constructing MLWFs, such as non-zero Chern numbers. These obstructions have been the basis of modern topological classification of crystalline insulators. In one dimension, however, there is a unique gauge that minimizes the spread functional of the WFs[77, 78].

### 2.1.3 Hybrid Wannier representation

A very useful tool for characterizing topological states is the hybrid Wannier (HW) representation. We start by selecting one crystallographic direction, say \( \hat{z} \), along which to
transform from \( k \) space to real space via the standard construction of one-dimensional (1D) maximally localized Wannier functions and their centers. This is done independently at each \( k_y \) in 2D, or at each \((k_x, k_y)\) in 3D. Thus, the HW functions are localized in real space along \( z \) but remain as extended Bloch-like functions in the orthogonal direction(s). The Wannier-center positions can be obtained from a parallel transport analysis performed independently for each \( k \)-point string encircling the BZ in the \( k_z \) direction. This kind of analysis also goes under the name of the “Wilson loop” \([43, 140, 141]\), the generalized non-Abelian Berry phases corresponding to the HW centers are frequently referred to as “Wilson-loop eigenvalues.”

To be explicit we consider a 3D insulator, we choose a reciprocal-lattice direction along which the HW representation will be constructed, and orient the Cartesian axes such that the corresponding real-space direction is \( z \). Given a gauge for the Bloch states \(|\psi_{nkk_z}\rangle\), the corresponding HW states are expressed as

\[
|h_{lnk}\rangle = \frac{1}{2\pi} \int_{-\pi/c}^{\pi/c} dk_z e^{-ik_zlc} |\psi_{nkk_z}\rangle.
\]

(2.18)

Here \( k = (k_x, k_y) \) is the wavevector in the perpendicular plane, i.e., in the projected two-dimensional Brillouin zone (2DBZ); \( c = 2\pi/b \) is the lattice constant along \( z \), with \( b \) the magnitude of the shortest reciprocal lattice vector along \( z \); \( l \) is an index that runs over unit cells along \( z \); and \( n \) runs over the \( J \) occupied bands in the insulator, which is the same as the number of HW functions at each \( k \) in one vertical unit cell.

Henceforth the multiband gauge of the Bloch wavefunctions is always taken such that the \(|h_{lnk}\rangle\) are maximally localized along \( z \) \([77]\). The HW wavefunctions \( h_{lnk}(r) = \langle r|h_{lnk}\rangle\) are thus Bloch-like and cell-periodic\(^2\) in the in-plane directions, while at each \( k \) they are the maximally localized Wannier functions of the effective 1D Hamiltonian \( H_k \) at that \( k \). Their centers

\[
z_{ln}(k) = \langle h_{lnk}|z|h_{lnk}\rangle
\]

(2.19)

form the Wannier “bands” (or “sheets”). These are periodic in real space along \( z \), \( z_{ln}(k) = z_{0n}(k) + lc \), as well as periodic in the in-plane reciprocal space, \( z_{ln}(k) =

\(^2\)By “cell-periodic” we mean that the \( e^{ik\cdot r} \) phase has been factored out, so that \( h_{lnk}(r) = h_{lnk}(r) + R \) for any in-plane lattice vector \( R \).
of \( z_{ln}(\mathbf{k} + \mathbf{G}) \), where \( \mathbf{G} \) is an in-plane reciprocal lattice vector. As noted earlier, this is essentially the same construction as that of the non-Abelian Wilson loop.

The Berry connection is defined in the HW representation as

\[
A_{ln,l'n'}^{\alpha} = \langle h_{ln} | i \partial_{\alpha} | h_{l'n'} \rangle, \tag{2.20}
\]

where \( \partial_{\alpha} = \partial / \partial k_{\alpha} \) with \( \alpha = \{x, y\} \). As a reminder, \( l \) and \( l' \) run over unit cells in the \( z \) direction, and \( n \) and \( n' \) label Wannier bands within the unit cell. Since there are \( J \) occupied bands in our insulator, \( n \) and \( n' \) run from 1 to \( J \). We shall also need to make use of the Wannier-band-diagonal Berry curvature

\[
\Omega_{ln,ln} = \partial_x A_{ln,ln}^y - \partial_y A_{ln,ln}^x
= -2 \text{Im} \langle \partial_x h_{ln} | \partial_y h_{ln} \rangle. \tag{2.21}
\]

Periodicity implies that

\[
A_{ln,l'n'}^{\alpha} = A_{0n,(l'-l)n'}^{\alpha}, \tag{2.22}
\]

and therefore \( \Omega_{ln,ln} = \Omega_{0n,0n} \).

Next, we discuss the gauge dependence of the quantities introduced above, as first presented in Taherinejad and Vanderbilt [111]. The Wannier bands are predetermined by the maximal localization procedure, so except in the case of degeneracy between Wannier bands, the most general gauge transformation that we need to consider is a Wannier-band-dependent phase twist

\[
| \tilde{h}_{ln} \rangle = e^{-i \beta_{ln}(k)} | h_{ln} \rangle. \tag{2.23}
\]

This leads to new Berry connections

\[
\tilde{A}_{ln,l'n'}^{\alpha} = e^{-i (\beta_{l'n'} - \beta_{ln})} (A_{ln,l'n'}^{\alpha} + \delta_{ln,l'n'} \partial_{\alpha} \beta_{ln}). \tag{2.24}
\]

The Berry curvature, on the other hand, is gauge invariant, which follows from the first line of Eq. (2.21) using \( \partial_x \partial_y \beta = \partial_y \partial_x \beta \).

---

3In the case that \( N \) Wannier bands are degenerate in some region of the 2DBZ, the present arguments can be generalized by considering a multiband gauge transformation involving a \( k \)-dependent \( N \times N \) unitary mixing among the degenerate Wannier bands; after tracing over any physical contribution over the degenerate bands, the same conclusions follow.
2.2 Berry phase in electronic structure

In this section we describe the physical manifestation of the mathematical quantities outlined in the previous section in the case of the electronic structure. Namely, the Berry phase manifests as the electric polarization in 1D and first Chern number as the quantum anomalous Hall effect in 2D. At the end of this section, we establish the physical connection between the two effects through the notion of adiabatic pumping and dimensional reduction. In this way, we built the necessary intuition for the next section in which we derive physical observables related to the higher dimensional analogs of the Berry phase and the first Chern number-the axion coupling and the second Chern number- which are the central aspect of this thesis.

2.2.1 Electric polarization

Electric polarization is one of the most fundamental physical quantities in solids. In almost all the traditional textbooks about electromagnetism, the electric polarization is defined as some kind of electric-dipole density. In crystalline solids, the polarization may be partitioned into two parts: one contributed from the bare nuclei and the localized core electrons $P_{\text{ion}}$, and the other from the valence electrons $P_{\text{el}}$. One would then assume that the knowledge of the bulk charge density would enable the calculation of the total bulk polarization. However, within periodic boundary conditions different choices of the unit cell can yield different values. Therefore it is natural to ask how do we define and calculate the bulk polarization. These answer came with “the modern theory of polarization” in the 1990s by Resta [98] King-Smith, and Vanderbilt [63].

To understand how bulk polarization is defined it is instructive to consider what is actually measured in experiments when one measures polarization. Indeed, experimentally polarization is not measured directly; instead one measures the current that flows when polarization is switched. This is illustrated in Fig. 2.3(a) where a ferroelectric is placed in a circuit where voltage oscillates over time and the current over time is measured. The resulting hysteresis loop in Fig. 2.3(b) has the electric field (from the voltage and sample thickness) on the horizontal axis and the time integrated current on
the vertical axis, now labeled $P$. So what is actually measured is the integrated current as the system switches between oppositely polarized variants, i.e.

$$\Delta P = \int_{i}^{f} J(t) dt$$

(2.25)

This quantity is not the polarization in any one state, but the change in polarization during switching. For an adiabatic evolution parametrized by $\lambda(t)$ we can change the integration variable from $t \rightarrow \lambda$ and using adiabatic perturbation theory \[124\] we find that the change in polarization from the state $\lambda_i$ to state $\lambda_f$ is (for simplicity we take the one-dimensional case)

$$\Delta P = -\frac{e}{2\pi} \int_{S} d\lambda dk \sum_{n}^{occ} \Omega_{\lambda k}^{n}$$

(2.26)

where the Berry curvature

$$\Omega_{\lambda k}^{n} = -2\text{Im} \langle \partial_{\lambda} u_{nk} | \partial_{k} u_{nk} \rangle$$

(2.27)

is defined over the $k - \lambda$ space as illustrated in Fig. 2.3(c).

Figure 2.3: (a) Ferroelectric of thickness $d$ in a capacitor configuration with voltage supply and ammeter indicated. (b) Typical hysteresis loop showing change of polarization $P$ with applied electric field. (c) The Berry curvature is defined over the $k - \lambda$ and it’s integral gives the change in polarization as one adiabatically evolves from $\lambda_i$ to $\lambda_f$. Adapted from Vanderbilt \[124\].

In addition we can use Berry connection to express $\Delta P$ in terms of only the end points using the notion of ”formal polarization”

$$\Delta P = P_{\lambda_f} - P_{\lambda_i}$$

$$P_{\lambda} = -\frac{e}{(2\pi)^{3/2}} \sum_{n} \int_{\text{BZ}} d\mathbf{k} A_{nk}(\lambda)$$

(2.28)
We see that the formal polarization is identified with the Berry phase. The polarization finds a more physical interpretation in the Wannier representation. If we construct Wannier functions from smooth periodic functions $\langle \mathbf{r} | w_{n\mathbf{k}} \rangle$. Then the Wannier charge centers(WCC) in the home unit cell are

$$r_n = \langle w_{n0} | \mathbf{r} | w_{n0} \rangle = \frac{V}{(2\pi)^3} \int_{\text{BZ}} d\mathbf{k} A_n \mathbf{k}$$

(2.29)

so that the polarization is simply the sum of WCC of occupied bands

$$\mathbf{P} = -\frac{e}{V} \sum_n r_n.$$  

(2.30)

In this way any physical change in the system, such as external fields or displacement of atoms, which changes polarization results in a shift in the position of WCC. In addition, an arbitrary change of gauge can shift individual WCCs in an arbitrary way but the sum of WCCs over the occupied manifold remains invariant modulo a lattice vector $\mathbf{R}$. Therefore we see that the formal polarization is only well defined modulo a quantum $e\mathbf{R}/V$, which is consistent with the fact only changes in polarization are physically meaningful.

### 2.2.2 Quantum anomalous Hall effect

Besides the electric polarization, another important application of the Berry phase in solids is the anomalous Hall effect. To see how the anomalous Hall effect arises consider the conductivity tensor $\sigma_{ab}$, that relates an applied electric field in the $b$ direction $E_b$ to the current density along the $a$ direction $J_a$

$$J_a = \sigma_{ab} E_b.$$  

(2.31)

The electric current density is given by the integral of the physical velocity of the electrons $\mathbf{v}_{n\mathbf{k}} = \langle u_{n\mathbf{k}} | \partial_{\mathbf{k}} H_{\mathbf{k}} | u_{n\mathbf{k}} \rangle$ weighed by their occupation function $f(\mathbf{k})$

$$\mathbf{J} = -e \sum_{n \in \text{occ}} \int_{\text{BZ}} \mathbf{v}_{n\mathbf{k}} f(\mathbf{k}).$$  

(2.32)

To linear response to a weak applied electric field one finds that in addition to the group velocity of the electron wave there is an additional anomalous velocity arising from the
presence of the external electric field

\[ \mathbf{v}_{nk} = \frac{\partial \epsilon_{nk}}{\partial \mathbf{k}} - \frac{e}{\hbar} \mathbf{E} \times \Omega_{nk} \]  

(2.33)

where \( \Omega_{nk} \) is the Berry curvature of the \( n \)th band. Physically, the Berry curvature results from the first order -in the rate of the change of the Hamiltonian- correction to the single particle eigenstates \( |u_{nk}\rangle \),

\[ |u_{nk}\rangle \rightarrow |u_{nk}\rangle - i\hbar \sum_{n' \neq n} \frac{|u_{n'k}\rangle \langle u_{n'k}| \partial_t u_{nk}}{\epsilon_n - \epsilon_{n'}}. \]  

(2.34)

To arrive at Eq. (2.33) we note that the electric field can enter the conversation through the vector potential \( \mathbf{A} = -\mathbf{E}t \) without breaking the crystal periodicity.\(^4\) In this case the only time dependence of the Hamiltonian enters through the gauge-invariant crystal momentum \( \mathbf{k} \) which in the presence of the electric field becomes \( \mathbf{k} \rightarrow \mathbf{k} - e/\hbar \mathbf{E}t \) so that

\[ \partial_t = -\frac{e}{\hbar} \mathbf{E} \cdot \partial \mathbf{k}. \]

For a two-dimensional system the Hall conductivity becomes

\[ \sigma_{xy} = \frac{e^2}{h} \int_{BZ} \frac{d^2 k}{2\pi} \sum_{n \in \text{occ}} \Omega_{nk} f(k), \]  

(2.35)

which is valid for both insulators and metals. For an insulator and at temperatures much lower than the energy gap, i.e. \( k_B T \ll E_g \), the group velocity term when integrated over the Brillouin zone gives zero contribution to the current density, while the anomalous velocity will have a non-zero contribution known as the Hall conductivity \( \sigma_H \),

\[ \sigma_H = \frac{e^2}{h} \int_{BZ} \frac{d^2 k}{2\pi} \sum_{n \in \text{occ}} \Omega_{nk}. \]  

(2.36)

Once again we encounter the situation where the Berry curvature is integrated over a closed manifold. Here \( \sigma_H \) is the first Chern number \( C \) in the units of \( e^2/h \), i.e.,

\[ \sigma_H = \frac{e^2}{h} C. \]  

(2.37)

Therefore at low temperatures the Hall conductivity is quantized for a two-dimensional band insulator of non-interacting electrons. Note that due to conservation of charge \( \partial_t \rho = -\partial_a J_a \), the QH response \( J_a = \sigma_H e^{ab} E_b \) implies another response equation \[93\],

\[ \partial_t \rho = -\sigma_H e^{ab} \partial_a E_b = \sigma_H \partial_t B. \]  

(2.38)

\(^4\)It could also enter through an inhomogeneous potential \( V(r) = \mathbf{E} \cdot r \).
Both the density and current response can be put together in a covariant way,

\[ J_\mu = \frac{C}{2\pi} \epsilon^{\mu\nu\tau} \partial_\nu A_\tau \quad (2.39) \]

where \( \mu, \nu, \tau = 0, 1, 2 \) are temporal and spatial indices.

In a field theoretic language the response in Eq. (2.39) can be described by an effective action \( S_{\text{eff}}^{(5)} \) known as the Chern-Simons field theory of the external field \( A_\mu \)

\[ S_{\text{eff}} = \frac{C}{4\pi} \int d^2 r \, dt \epsilon^{\mu\nu\tau} A_\mu \partial_\nu A_\tau. \quad (2.40) \]

The covariant response Eq. (2.39) is then simply extracted from the equation of motion by

\[ J_\mu = \frac{\delta S_{\text{eff}}}{\delta A_\mu} \quad (2.41) \]

with the coefficient \( C \) being material dependent and coming from response theory. Such an effective action is topologically invariant, in agreement with the topological nature of the first Chern number.

Historically the quantization of the Hall conductivity in a crystal was first shown by Thouless et al. [115] for the integer quantum Hall effect (IQHE). Haldane [47] then showed that a non-zero Chern number does not require an external magnetic field by constructing a tight-binding model on a honeycomb lattice which displayed the quantum anomalous Hall effect with zero net flux per unit cell.

We should note that Eq. (2.35) is valid for metals as well as for insulators. In this case the Hall conductivity \( \sigma_{xy} \) in not quantized and can be expressed as a property of the Fermi surface (at low temperature)

\[ \sigma_{xy} = \frac{e^2}{\hbar} \int_{\text{BZ}} \frac{d^2 k}{2\pi} \sum_{n \in \text{occ}} \Omega_{nk} f(k) \]

\[ = \frac{e^2}{\hbar} \int_{\text{BZ}} \frac{d^2 k}{2\pi} \sum_{n \in \text{occ}} A_{nk}^x \frac{\partial f(k)}{\partial k_y} - A_{nk}^y \frac{\partial f(k)}{\partial k_x} \]

\[ = \frac{e^2}{\hbar} \int_{\text{BZ}} \mathbf{k} \cdot \mathbf{A}_{nk} \delta(E - E_F) = \frac{e^2}{\hbar} \phi(C_F) \quad (2.42) \]

---

In our case the effective action \( S_{\text{eff}}[A_\mu] \) is obtained from the full action \( S[A_\mu, \bar{\psi}, \psi] \) describing the interaction between matter fields \( \bar{\psi}, \psi \) and an external gauge field \( A_\mu \) by integrating out the matter fields. That is \( e^{iS_{\text{eff}}[A_\mu]} = \int D\bar{\psi} D\psi e^{iS[A_\mu, \bar{\psi}, \psi]} \).
where $C_F$ refer to the Fermi loop(s) and $\phi(C_F)$ is the Berry phase around $C_F$. Note that we used the fact that at zero temperature the partial derivative of $f(k)$ behaves as a $\delta-$function picking-up the states at the Fermi level. This expression will become important when we discuss the bulk-boundary correspondence of topological insulators.

### 2.2.3 Adiabatic pumping

Polarization in 1D and the quantum anomalous effect in 2D are intriguingly related. To see this relationship let us consider an 1D chain and a parametric variation of the ground state over a closed loop through a parameter $\lambda$ with $\lambda_i = \lambda_f$. If we want to calculate the polarization, then the manifold $S$, in Eq. (2.26), we are integrating the Berry curvature over is a torus. Since a torus is a closed manifold the Chern theorem applies and it tells us that

$$
\Delta P = -e \sum_n C_n = -eC.
$$

(2.43)

where $C_n$ are the Chern number of individual bands

$$
C_n = \phi_n(\lambda_f) - \phi_n(\lambda_i).
$$

(2.44)

and $C$ is the total Chern number of the occupied band subspace along this path. The physical interpretation is that exactly $C$ electrons have been pumped one unit cell to the right along the chain during the cyclic adiabatic evolution. This can be easily seen by reminding ourselves that the Wannier centers and Berry phases are related by $\bar{x}_n = a/2\pi \phi_n$. This quantized transport through an adiabatic evolution was first described by Thouless [114] and is known as a Thouless pump. An illustration of a Thouless pump with $C = 1$ is shown in Fig. 2.4(a).

It is instructive to flip the question. Starting with a 2D Bloch Hamiltonian $H(k_x,k_y)$ coupled to an external field $A_\mu$, $\mu = 0, 1, 2$ we construct the family of 1D systems $H_\lambda(k_x)$ defined by identifying one of the k-space directions with the adiabatic parameter, i.e. $k_y + e^2/hA_y(t) = \lambda(t)/a$. Then the effective action of the 1D system is by construction the same as in Eq. (2.40) but there are no integrals along $y$

$$
S_{\text{eff}} = \frac{G(\lambda)}{4\pi} \int dx dt \lambda \epsilon^{\mu\nu} \partial_\mu A_\nu
$$

(2.45)
and similar the pumping coefficient $G$

$$G(\lambda) = -\int_{BZ} \frac{dk_x}{2\pi} \Omega_{nk_x} = \int_{BZ} \frac{dk_x}{2\pi} (\partial_{k_x} A_\lambda - \partial_\lambda A_{k_x}).$$

(2.46)

The pumping coefficient $G(\lambda)$ satisfies the sum rule

$$C = \oint d\lambda G(\lambda)$$

(2.47)

and if we choose a proper gauge where $A_\lambda$ is always single valued, the pumping coefficient becomes

$$G(\lambda) = \partial_\lambda \left( \int_{BZ} \frac{dk_x}{2\pi} A_{k_x} \right) = \partial_\lambda P(\lambda),$$

(2.48)

where in the second equality we have identify the integral with the charge polarization of the 1D system. This theoretical construction known as “dimensional reduction” provides us with the response of an 1D system under the influence of an adiabatic field $\lambda(t)$. In light of the above, the effective action becomes

$$S_{\text{eff}} = \frac{\partial_\lambda P}{4\pi} \int dx dt \lambda \epsilon^{\mu\nu} \partial_\mu A_\nu$$

$$= -\frac{1}{4\pi} \int dx dt (\partial_\lambda P \partial_\mu \lambda) \epsilon^{\mu\nu} A_\nu$$

$$= -\frac{1}{4\pi} \int dx dt (\partial_\mu P) \epsilon^{\mu\nu} A_\nu$$

(2.49)

$$= \frac{1}{4\pi} \int dx dt P \epsilon^{\mu\nu} \partial_\mu A_\nu.$$
Note that there is no more dependence on the adiabatic parameter $\lambda$. As Qi et al. [93] pointed out the above considerations hold even when the adiabatic parameter has a smooth spatial dependence, i.e. when $\lambda(t) \to \lambda(x, t)$, and the current response can be extracted from the equation of motion Eq. (2.41) which gives

$$j_\mu = -\epsilon^{\mu\nu} \partial_\nu P[\lambda(x, t)]$$

which spelled out tells us that

$$j_x = \partial_t P(\lambda)$$
$$\rho = -\partial_x P(\lambda).$$

This means, as expected, the current and density response of the system do not depend on the parametrization.

Turning back to the quantum anomalous Hall effect where a constant electric field $E_y$ is applied at a 2D insulator. The electric field can enter through the vector potential, in which case the gauge-invariant crystal momentum $k$ becomes

$$k_x = k_x|_{E=0}$$
$$k_y(t) = k_y|_{E=0} + \frac{e}{\hbar} A_y(t) = k_y - \frac{e}{\hbar} E_y t.$$  

Finally we identify $\lambda(t)/a \to k_y - \frac{e}{\hbar} E_y t$ as the adiabatic parameter. Then the current response after an adiabatic cycle $\lambda = 0 \to 2\pi$ is given by

$$J_x = \int_0^{2\pi} d\lambda J_x(\lambda) = -\frac{eE_y}{\hbar} \int_0^{2\pi} d\lambda \frac{\partial P_x(\lambda)}{\partial \lambda} = -\frac{eE_y}{\hbar} [P_x(2\pi) - P_x(0)].$$

Consequently, the change $\Delta P_x = P_x(2\pi) - P_x(0)$ through a period of adiabatic evolution is an integer equal to $-C$ and corresponds to the charge pumped through the system.

In summary the current $J_x$ in response to an electric field $E_y$ in the quantum Hall effect is equivalent to a charge pump of the effective 1D system obtained from the dimensional reduction procedure described in Sec. 2.2.3.

### 2.3 Axion coupling in electronic structure

In this section we discuss the physics of the axion magnetoelectric coupling. By drawing analogies between the relationship of the first Chern number and the Berry phase we
explain how the axion coupling describes the responses of a 3D insulator in the presence of external electromagnetic fields. We then consider the symmetries that quantize the axion coupling to define a $\mathbb{Z}_2$ topological classification and describe the properties of the topologically non-trivial phase. We finish this section with a discussion of the recent advances and open questions in the field.

2.3.1 From the second Chern number to the axion coupling

In the previous section we discussed the idea of a Thouless pump and dimensional reduction. Since Chern numbers are well defined in every even dimension, it is natural to ask if they can be associated with any physical observable. Namely, the second Chern number $C^{(2)}$ is an integer invariant of 4D closed manifolds and we can ask if it describes a response property of 4D band insulators. This was first done in the seminal paper of Qi et al. [93], who took considered the effective field theory of a 4D band insulator in the presence of a $U(1)$ gauge field.

Remarkably, the second Chern number appears as a nonlinear-response coefficient of 4D band insulators in the presence of an external $U(1)$ gauge field, which is in exact analogy with the first Chern number $C$ as the Hall conductance of a 2D system. In this case the effective action, in natural units, becomes

$$S_{\text{eff}} = \frac{C^{(2)}}{24\pi^2} \int d^4r \, dt \epsilon^{\mu\nu\rho\sigma\tau} A_\mu \partial_\nu A_\rho \partial_\sigma A_\tau$$

(2.54)

where $A_\mu$ is the external gauge field and $C^{(2)}$ is an integer invariant calculated from the ground state of the 4D insulator,

$$C^{(2)} = \frac{1}{32\pi^2} \int_{BZ} d^4k \epsilon^{\mu\nu\sigma\tau} \text{Tr} [\Omega_{\mu\nu} \Omega_{\sigma\tau}]$$

(2.55)

where $\mu, \nu, \sigma, \tau$ run over $k_x, k_y, k_z, k_w$ and $\Omega_{\mu\nu}$ is now the non-Abelian Berry curvature given by

$$\Omega_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu - [A_\mu, A_\nu].$$

(2.56)

Then the current response is extracted from the equation of motion giving

$$j^{4D}_\mu = \frac{C^{(2)}}{8\pi^2} \epsilon^{\mu\nu\rho\sigma\tau} \partial_\nu A_\rho \partial_\sigma A_\tau$$

(2.57)
which describes a non-linear response to the external field $A_\mu$, in contrast with the linear response of 2D band insulators in Eq. (2.39).

One would naively think that this is interesting only from a mathematical perspective since we are not living in 4D. However, this is not the case. Looking back to the previous section we explained how that the QHE describes a family of 1D systems -obtained from dimensional reduction procedure- that pump $-C$ electrons after an adiabatic cycle. Therefore, we can ask what quantity, analogous to 1D polarization, can we pump in 3D to give us the non-linear response in Eq. (2.57). The answer came from Qi et al. [93] who identified the magnetoelectric polarization $P_3$ (which we will also refer to as the axion coupling $\theta_{cs}$) as the 3D analog of the electric polarization in 1D.

Similar to the QHE case, in Sec. 2.2.3 we consider the Hamiltonian of a 4D insulator coupled to $U(1)$ gauge field. We also consider a gauge where the vector potential $A_\mu$ is translationally invariant in the 4th direction so for periodic boundary conditions $k_4$ is a good quantum number. We then associate the gauge invariant momentum with an adiabatic parameter field $\lambda(r,t)$

$$\lambda(r,t)/a = k_4 + e^2/h A_4(r,t) \quad (2.58)$$

like in Eq. (2.52). In this way the 4D insulator coupled to the gauge vector field $A_\mu$, $\mu = 0, 1, 2, 3, 4$ becomes a 3D insulator coupled to the gauge vector field $A_\nu$, $\nu = 0, 1, 2, 3$ and the adiabatic scalar field $\lambda(r,t)$. To continue we expand the effective action around a constant field configuration $A_\nu(r,t) = 0$ and $\lambda(r,t) = \bar{\lambda}$. Then by construction the effective action $S_{\text{axion}}$ is identical to the 4D case but without the integration over the $k_4$ integral, that is

$$S_{\text{axion}} = \frac{G_3(\bar{\lambda})}{4\pi} \int d^3r \, dt \epsilon^{\mu \nu \rho \sigma} \delta \lambda \partial_\mu A_\nu \partial_\rho A_\sigma. \quad (2.59)$$

Here $\delta \lambda(r,t) = \lambda(r,t) - \bar{\lambda}$ encodes the small variations from the constant background field and $G_3(\bar{\lambda})$ is the pumping coefficient of the 3D system in analogy with Eq. (2.46) in the 1D case and is given by

$$G_3(\bar{\lambda}) = \frac{1}{8\pi^2} \int_{\text{BZ}} d^3k \epsilon^{ijk} \text{Tr} \left[ \Omega_{\lambda i} \Omega_{jk} \right], \quad (2.60)$$
in which the Berry curvature is defined in the 4D space \((k_x, k_y, k_z, \bar{\lambda})\). In analogy with the pumping coefficient \(G(\bar{\lambda})\) in Eq. (2.47), \(G_3\) satisfies the sum rule

\[
C^{(2)} = \int d\bar{\lambda} G_3(\bar{\lambda}).
\]  

(2.61)

Recall that \(G(\lambda)\) was expressed as \(\partial_\lambda P(\lambda)\) where \(P(\lambda)\) is the polarization of the 1D system. This is true for \(G_3(\bar{\lambda})\) which can be expressed as \(G_3(\bar{\lambda}) = \partial_\lambda P_3(\bar{\lambda})\) so that the magnetoelectric polarization \(P_3\) is the analog of 1D polarization \(P\) in 3D. To find the expression for \(P_3\) we use the fact that the integrand of Eq. (2.60) can be written as a total derivative

\[
G_3(\bar{\lambda}) = \int_{BZ} d^3k \partial_\mu K_\mu
\]

(2.62)

with \(\mu, \nu, \rho, \sigma\) running over \(x, y, z, \bar{\lambda}\). Here \(K_\mu\) is known as the non-Abelian Chern-Simons 3-form and mathematically it represents a 4D vector in the parameter space \((k_x, k_y, k_z, \bar{\lambda})\). Next, we are allowed to choose a gauge where \(K_x, K_y, K_z\)-but not \(K_\bar{\lambda}\)-are single valued functions so that the pumping coefficient becomes

\[
G_3(\bar{\lambda}) = \int d^3k \partial_\lambda K_\bar{\lambda} = \partial_\lambda P_3(\bar{\lambda})
\]  

(2.63)

with the magnetoelectric polarizability \(P_3\) given by the integral of the Chern-Simons 3-form

\[
P_3(\lambda) = \frac{1}{16\pi^2} \int d^3k \epsilon^{\lambda\alpha\beta\gamma} \text{Tr}[(\Omega_{\alpha\beta} - \frac{i}{3}[A_\alpha, A_\beta]) \cdot A_\gamma].
\]  

(2.64)

where \(\alpha, \beta, \gamma\) runs only over \(x, y, z\). Having found an expression for \(P_3\) we return back to the axion action Eq. (2.59) to replace \(G_3\) and perform an integration by parts to find

\[
S_{\text{axion}} = \frac{1}{4\pi} \int d^3r dt \epsilon^{\mu\nu\rho\sigma} A_\mu(\partial_\lambda P_3 \partial_\nu \delta \lambda) \partial_\rho A_\sigma
\]

(2.65)

where \(\partial_\lambda P_3 \partial_\nu \delta \lambda = \partial_\nu P_3\). We then undo the first integration by parts -by performing another one- to remove the derivative from \(P_3\)

\[
S_{\text{axion}} = \frac{1}{4\pi} \int d^3r dt \epsilon^{\mu\nu\rho\sigma} P_3 \partial_\nu (A_\mu \partial_\rho A_\sigma)
\]  

(2.66)
which remarkably results in an action that is independent of adiabatic parameter $\lambda$.

Lastly, we express everything in terms of the electric and magnetic fields and restore the units to arrive at our final axion action

$$S_{\text{axion}} = \frac{e^2}{2\pi \hbar} \int d^3 r dt P_3 E \cdot B \quad (2.67)$$

From now one instead of $P_3$ we will be using the equivalent axion magnetoelastic coupling $\theta_{cs}$ or simply axion coupling and the two are simply related by $\theta_{cs} = P_3/2\pi$. Simplifying Eq. (2.64) gives as an expression for the axion coupling in the Bloch representation

$$\theta_{cs} = -\frac{1}{4\pi} \int_{\text{BZ}} d^3 k \epsilon^{\alpha\beta\gamma} \text{Tr} \left[ A_\alpha \partial_\beta A_\gamma - i \frac{2}{3} A_\alpha A_\beta A_\gamma \right].$$  

(2.68)

Being a pseudoscalar characterizing the ground state of a 3D insulator, one might expect the axion coupling $\theta_{cs}$ to be gauge invariant, but this is not the case. In fact, a gauge transformation, that is, a $k$-dependent unitary transformation $U_{mn}(k)$ that mixes occupied bands, can cause $\theta_{cs} \rightarrow \theta_{cs} + 2\pi N$ for some integer $N$. To see this note that the Berry connection transforms in the following way

$$A_\alpha \rightarrow U \dagger A_\alpha U + U \dagger i \partial_\alpha U$$

(2.69)

so that the axion coupling changes by

$$\Delta \theta_{cs} = \frac{1}{12\pi} \int_{\text{BZ}} d^3 k (U \dagger \partial_\alpha U)(U \dagger \partial_\beta U)(U \dagger \partial_\gamma U)$$

(2.70)

Since the above integral corresponds to the homotopy invariant $\pi_3(SU(N)) = \mathbb{Z}$ for $N \geq 2$, which classifies mappings from the 3-sphere (or 3-torus) to SU($N$), $\Delta \theta_{cs} = 2\pi n$ with $n \in \mathbb{Z}$ for two or more occupied bands.\(^6\) This implies that the only well defined part of the bulk axion coupling lives on a unit circle and the axion coupling should be thought of as a phase angle.

\(^6\)For a single occupied band $\pi_3(U(1)) = 1$ so $\Delta \theta_{cs} = 0$ and this leads to a refined classification known as a Hopf insulator [81].
2.3.2 Axion electrodynamics

In summary, we have found an effective action that generates the so-called axion electrodynamics

\[ S_{\text{axion}} = \frac{e^2}{2\pi \hbar} \int d^3 r dt \frac{\theta_{cs}}{2\pi} E \cdot B \]  

(2.71)

with the axion coupling \( \theta_{cs} \) being a material-dependent parameter determined by the ground state wave function as in Eq. (2.68). Note that Eq. (2.68) defines \( \theta_{cs} \) as a number, but as long as we are careful not to probe small distances, e.g. by setting a cut-off for small lengths, we can upgrade it to a scalar field \( \theta_{cs}(r,t) \). Since \( E \cdot B = \partial_\mu (A_\nu \partial_\rho A_\sigma) \), the Lagrangian can be written as a total derivative, and therefore it does not modify the equations of motion. However, the presence of the axion field can have profound consequences at surfaces and interfaces, where gradients in \( \theta_{cs}(r,t) \) appear. Calculating the equation of motion Eq. (2.41) for the axion action Eq. (2.71) we find

\[ J_i = \epsilon^{ijk} (\partial_j \alpha_{cs}(r,t)) E_k + (\partial_t \alpha_{cs}(r,t)) B_i \]  

(2.72)

where

\[ \alpha_{cs}(r,t) = \frac{e^2 \theta_{cs}(r,t)}{\hbar} \]  

(2.73)

is the dimensionful axion coupling. For uniform and static electric and magnetic fields the axion coupling describes the Chern-Simons contribution to the magnetoelectric tensor \( \alpha_{ij} = (\partial P_i/\partial B_j)_E = (\partial M_j/\partial E_i)_B \). In this case we can identify them with the magnetization and polarization currents

\[ J_i^M = \epsilon^{ijk} \partial_j M_k = \epsilon^{ijk} \partial_j (\alpha_{cs}(r,t) E_k) \]

\[ J_i^P = \partial_t P_i = \partial_t (\alpha_{cs}(r,t) B_i) \]  

(2.74)

so that the response Eq. (2.41) is interpreted as a change in the magnetization \( \mathbf{M} \) and polarization \( \mathbf{P} \) which are now given by

\[ \mathbf{M} = \mathbf{M}_0 + \alpha_{cs} \mathbf{B} \]  

\[ \mathbf{P} = \mathbf{P}_0 + \alpha_{cs} \mathbf{E} \]  

(2.75)

More generally the first term in Eq. (2.41) describes the anomalous Hall effect while the second describes the chiral magnetic effect.
2.3.3 Topological insulators

Quantizing symmetries — The $\mathbf{E} \cdot \mathbf{B}$ term in the axion action Eq. (2.71) is odd under time reversal ($\mathcal{T}$) and inversion ($\mathcal{I}$). For a bulk 3D crystal that respects either of these symmetries this implies $\theta_{cs} = -\theta_{cs}$, and since $\theta_{cs}$ is a phase angle, it can only take the values 0 or $\pi$. A system with $\mathcal{T}$ and $\theta_{cs} = \pi$ is usually denoted as a strong topological insulator (TI), while in the absence of $\mathcal{T}$ an inversion symmetric crystal is usually called an axion insulator. We will later discuss the properties of strong TIs and axion insulators, but before we do that we want to enumerate all quantizing symmetries. Indeed, many other symmetries can quantize $\theta_{cs}$, such that $\theta_{cs}/\pi = 0$ or 1 defines an “axion $\mathbb{Z}_2$ index.” We use the term “axion-odd insulators” to refer to systems in which the nontrivial $\mathbb{Z}_2$ index is protected by one of these symmetries, with TR-protected strong TIs and inversion-protected axion insulators as special cases.

To decide whether $\theta_{cs}$ is quantized to 0 or $\pi$ by a set of crystal symmetries – i.e., whether the axion $\mathbb{Z}_2$ index is protected – we can just look at whether there are any elements in the magnetic point group that reverse the sign of $\theta_{cs}$. We shall call these the “axion-odd” symmetries. Since an n-fold rotation $C_n$ or a fractional translation $\tau_{p/q}$ \textsuperscript{7} do not affect the sign of $\theta_{cs}$, all axion-odd symmetries can be written as

\[
(\mathcal{T} \text{ or } \mathcal{I}) \ast C_n \ast \tau_{p/q}
\]

where (*) refers to composition. Examples of axion-odd symmetries include mirrors, glide-mirrors, proper rotations composed with TR, and improper rotations not composed with TR.

If one or more of these symmetries is present in the magnetic point group, then $\theta_{cs}$ is quantized to be 0 or $\pi$, i.e., the $\mathbb{Z}_2$ index exists. In many cases the magnetic point group may contain several axion-odd symmetry operations, but any one of them is enough to quantize $\theta_{cs}$.

Surface response — To see what is so special about topological insulators we need to consider their surfaces. In the presence of a uniform external electric field, e.g. along

\textsuperscript{7} $\tau_{p/q}$ is a fraction with $0 < p/q < 1$. 

Figure 2.5: (a) The topological magnetoelectric effect manifests as half-integer surface anomalous Hall conductivity. Adapted from Vanderbilt [124]. (b) A strong topological insulator is required to have an odd number of Dirac cones on every surface. (c) The gapped surfaces of an axion insulator exhibit a half quantized anomalous Hall conductivity.

\[ \hat{x} \text{ as in Fig. 2.5(a), the axion electrodynamics response Eq. (2.74) becomes} \]

\[ J_i^M = \epsilon^{ijk} \partial_j M_k = \frac{e^2}{\hbar} \epsilon^{ijk} \partial_j (\theta_{\text{cs}}(r,t)) E_k. \] \hspace{1cm} (2.77)

At the boundary between a topological insulator and trivial insulator, the axion coupling changes from \( \pi \) to 0. Due to the spatial derivative in Eq. (2.77) there is a non-zero response localized at the surface which contributes to the Hall current. The surface anomalous Hall conductivity (AHC) due to \( \theta_{\text{cs}} \) is then

\[ \sigma_{\text{surf}}^{\text{AHC}} = -\frac{e^2}{\hbar} \frac{\theta_{\text{cs}}}{2\pi} \text{ mod } e^2/h, \] \hspace{1cm} (2.78)

In this exact relation, the ambiguity modulo 2\( \pi \) in \( \theta_{\text{cs}} \) is consistent with a freedom to prepare insulating surfaces with values of \( \sigma_{\text{surf}}^{\text{AHC}} \) differing by the quantum \( e^2/h \), e.g., by changing the Chern number of some surface bands, or of adding or deleting a surface layer with a nonzero Chern number. If all surfaces adopt the same branch choice – i.e., the same value of \( \sigma_{\text{surf}}^{\text{AHC}} \) – then the sample as a whole exhibits a true magnetoelectric response of \(-\alpha_{\text{CS}}\), where the quantized part of the response has been absorbed into the branch choice for \( \alpha_{\text{CS}} \). This phenomenon is again a higher-dimensional analog of the modern theory of electric polarization [63], where the 2\( \pi \) ambiguity of the Berry phase reflects the inability to define the bulk polarization, or to predict the bound charge density of an insulating surface, except modulo a quantum.
Importantly, there is another contribution to the surface anomalous Hall conductivity that we need to take into account. As we discussed in the previous section the Hall conductivity of a 2D metal is given by the Berry phase of the states at the Fermi level, see Eq. (2.42). In contrast to the bulk contribution coming from $\theta_{cs}$, this is a purely surface property and so it does not suffer from any ambiguity. In summary, the surface response of a topological insulator can be written as

$$\sigma_{\text{AHC}}^{\text{surf}} = -\frac{e^2}{h} \theta_{cs} - \phi \mod \frac{e^2}{h}. \quad (2.79)$$

where the extra $\phi$ term is the Berry phase taken along the surface Fermi loop.

**Strong topological insulators** — The subset of topological insulators (TI) with the axion-odd symmetry being $T$ are known as strong TIs. Since $T$ symmetry forces $\sigma_{\text{AHC}}^{\text{surf}} = 0$ the Berry phase $\phi$ contribution has to cancel the $\theta_{cs} = \pi$ contribution on any surface. This Berry phase $\phi = \pi$ then manifests as an odd number of Dirac cones on any surface.

Being the only topological insulators whose axion-odd symmetry does not contain any crystalline symmetry, strong TIs occupy a special place among the axion-odd insulators due to their unique properties. In fact, the discovery of the $Z_2$ index for 3D $T$-symmetric insulators preceded the discovery of the axion coupling [40, 80]. In passing, we also note that the strong $Z_2$ classification is tabulated in the tenfold way (class AII). In this classification, ten symmetry groups are considered based on the presence or absence of time-reversal and charge conjugation symmetry. These groups are especially important in that they remain well-defined even in the presence of disorder.

Historically, the story of strong TIs is a triumph of condensed matter physics in how beautifully theory and experiment worked together. Indeed, only two years after the theoretical work of Fu and Kane [40] and Moore and Balents [80] on the topological classification of 3D $T$-symmetric insulators, Xia et al. [143] identified Bi$_2$Se$_3$ as a strong TI with a single Dirac cone centered at the $\Gamma$ point in the 2D surface BZ. Simultaneously, a pioneering paper by Zhang et al. [152] provided a detailed understanding of the role of spin-orbit coupling in driving the band inversion at the $\Gamma$ point that leads to the topological state in this material.

**Magnetic topological insulators** — We have seen that time-reversal symmetry $T$
forces the surface anomalous Hall conductivity of a topological insulator to vanish. However, if $T$ is locally broken on a surface, then the bulk contribution to $\sigma_{\text{AHC}}^{\text{surf}}$ coming from $\theta_{\text{cs}} = \pi$ is unaffected while the surface Dirac cone can be gapped out. This realization came soon after the discovery of strong TIs and led scientists to ferromagnetically dope thin films of strong TIs. The Dirac cones on the top and bottom surfaces are then gapped out and the surfaces exhibit a half-integer anomalous Hall conductivity. In the thin film limit the contributions from the two surfaces are added and the side surfaces become the side edges. Finally, since the magnetic dopants order ferromagnetically, the contributions have the same sign in a global sense. In this way the 2D quantum anomalous Hall effect was experimentally observed for the first time [20]. Unfortunately the inhomogeneity of the magnetic dopants leads to inevitable disorder [69], and as a result the quantized response is observed at much lower temperatures than the magnetic gap and Curie temperature allow, see Fig. 2.6(a).

More recently, there have been efforts to find intrinsically magnetic TIs. These systems have the potential to operate at higher temperatures compared to doped TIs, due in part to the absence of magnetic-impurity disorder. Since bulk $T$ symmetry is broken, at least one axion-odd crystalline symmetry has to be present. Then depending on the surface termination the axion-odd symmetry may force a vanishing AHC, in which case an odd number of Dirac cones will appear, or it will not impose any constraints, in which case a gapped surface will exhibit a half-integer AHC. In the former case the axion-odd symmetry is respected at the surface while in the latter it is not.

The first example of an intrinsically magnetic TI was experimentally observed in MnBi$_2$Te$_4$ [87]. This compound has a van der Waals structure with the layers stacked along the $\hat{z}$ direction. The magnetic ordering is A-type AFM, i.e., with magnetization uniform in-plane but alternating from plane to plane along the stacking direction as illustrated in Fig. 2.6(b). MnBi$_2$Te$_4$ has both inversion $I$ and $T$ composed with a half translation along $\hat{z}$ ($T \ast \tau_{1/2}$) so that there are more than one axion-odd symmetries. However, because inversion is never a good surface symmetry, it does not impose any

---

8Formally this means the symmetry belongs to the surface group which is defined by the symmetry elements of the same crystal in a semi-infinite geometry with a single surface.
constrains on the surfaces. In contrast, surfaces like (100) or (010) are forced to have vanishing AHC due to $\mathcal{T} \ast \tau_{1/2}$.

Figure 2.6: (a) Chang et al. [20] showed the first realization of the quantum anomalous Hall effect in Ch doped thin film of topological insulator. At very low temperature the Dirac cones gap out resulting in an integer-quantized anomalous Hall conductivity when the Fermi level is tuned in the insulating gap. (b) Thin films of the antiferromagnetic topological insulator MnBi$_2$Te$_4$ exhibit the integer-quantized anomalous Hall effect at elevated temperatures. Adapted from Deng et al. [30].

**Symmetry indicators** — Even though the expression for $\theta_{cs}$ in Eq. (2.68) is elegantly written in its integral form, applying it is computationally expensive and somewhat problematic[27]. Fortunately, things are greatly simplified if one has a knowledge of the parities at the time-reversal invariant momenta (TRIM). As shown by Fu and Kane [40], a counting of the parities at the TRIM determines the $\mathbb{Z}_2$ invariant when $\mathcal{T}$ and $\mathcal{I}$ are both present. Later the parity-counting rule was generalized to the inversion-only case. [53, 121] This rule states that if the material is insulating and is not a Chern insulator,$^9$ then $\theta_{cs} = \pi$ if and only if the total number of odd-parity states (NOPS) at the TRIM is twice an odd integer.

---

$^9$In general, a three-dimensional insulator is characterized by three integer Chern indices; if any of these are nonzero, the system is a Chern insulator, and it displays a quantized anomalous Hall conductivity.
In recent years, generalizations of the Fu-Kane criterion take into account the full space group of a crystal. Namely, using constraints on band representations between high-symmetry points in the Brillouin zone, these elaborate schemes compare the real space structure of crystals with the momentum space structure, i.e. band structure, to construct indicators of topological non-triviality. These constructions known as symmetry indicators [92] or elementary band representations [15], have enabled high-throughput prediction of topological materials [128].
Chapter 3
Surfaces of axion insulators

In this chapter we explore various topological phenomena at the surface of an axion insulator, many of which cannot be obtained from a strong TI with broken $T$ on its surface. In doing so we hope to raise the awareness of the community regarding this promising class of materials, in the expectation that they may soon be realized and their unique surface properties explored. We investigate the conditions under which the surface magnetic point group forces the surface AHC to vanish, and when it does not, what factors decide the sign of the surface AHC. We find that the answers depend on the bulk and surface symmetry, the specific surface termination, and the magnetic configuration of the surface, yielding a plethora of possibilities to explore and manipulate.

Specifically, we carry out these investigations in the context of a model for the $R_2\text{Ir}_2\text{O}_7$ pyrochlore iridates ($R$ is typically a rare earth). Many members of this class exhibit a low-temperature insulating phase with AIAO antiferromagnetic order, although some also display FM or other magnetic orderings [31, 32, 104, 118, 154]. The AIAO pyrochlores were proposed early on by Wan et al. [130] to be in an axion-insulator phase for some range of values of the on-site Hubbard $U$. However, they become trivial insulators for sufficiently large values of $U$, and later studies have led to a consensus that these systems are in fact in the trivial phase (see, e.g., Zhang et al. [153]).

Nevertheless, motivated by the interest in this material system, we have adopted a minimal tight-binding model of the magnetic sites in a pyrochlore iridate, but with weakened magnetic exchange splitting, as a platform for investigating hypothetical axion-insulator materials. Our model consists of spinors on the Ir sites; these form a simple pyrochlore lattice, and therefore may also be relevant to some spinels. The Hamiltonian includes on-site Zeeman splitting terms representing the effects of the
magnetic order, as well as spin-dependent and independent nearest-neighbor hoppings. The phase diagram of our model exhibits a trivial insulator phase for large Zeeman splitting, and then passes through a Weyl semimetal phase before entering an axion-insulator phase as the Zeeman splitting is reduced. We explore the behavior of the bulk and surface properties of our model primarily in this axion-insulator region of the phase diagram, focusing mainly on the case of AIAO magnetic order where the symmetry is highest, but considering FM and other orderings as well. In particular, we make use of a recently introduced numerical approach [95], to derive an expression that allows for an explicit calculation of the magnitude and sign of the surface AHC. In turn, this tool is used to demonstrate the half-integer nature of the surface AHC and to reveal how the surface structural and magnetic configuration determines its sign.

3.1 Calculation of the surface anomalous Hall conductivity

For the calculation of the surface anomalous Hall conductivity (AHC), we implement a recently proposed approach [95] based in part on previous developments [9, 35, 76] showing that the Chern-Simons (CS) contribution to the AHC can be expressed as a local, real space property. In this framework, one defines a local Chern-number density

$$C(r) = -2\pi \text{Im} \langle r | P r Q \times Q r P | r \rangle$$

(3.1)

from which the local CS contribution to the AHC can be obtained via $\sigma^{\text{CS}}(r) = (e^2/h)C(r)$. Here $P$ and $Q$ are the projection operators onto the occupied and unoccupied subspaces, respectively. In contrast with the usual reciprocal-space integral expression, the real space one can be used to study bounded systems such as finite crystallines or surface slabs. This is exactly what we need to calculate the surface AHC and determine its sign.

We apply this theory in the context of a surface slab geometry. The unit cell is small in the in-plane direction, with primitive lattice vectors $a_1$ and $a_2$ and cell area $A = |a_1 \times a_2|$, but it extends through the entire thickness of the slab in the vertical $z$ direction, including top and bottom surfaces. The Hamiltonian eigenstates in the valence ($v$) and conduction ($c$) bands are $\psi_{v\mathbf{k}}(r)$ and $\psi_{c\mathbf{k}}(r)$, respectively, with $\mathbf{k} =$
\((k_x, k_y)\). The valence and conduction projectors are then \(P = (1/N_k) \sum_{vk} |\psi_{vk}\rangle \langle \psi_{vk}| \) and \(Q = (1/N_k) \sum_{ck} |\psi_{ck}\rangle \langle \psi_{ck}| \), where \(N_k\) is the number of \(k\)-points in the 2D Brillouin zone mesh. Plugging these expressions into Eq. (3.1), we find

\[
C_z(r) = -\text{Im} \frac{4\pi}{N_k} \sum_{vk} \sum_{vvc} \psi_{vk}(r) X_{vcck} Y_{vc'ck}^\dagger \psi_{vc'k}(r) \tag{3.2}
\]

where

\[
X_{vcck} = \langle \psi_{vk} | x | \psi_{ck} \rangle = \frac{\langle \psi_{vk} | i\hbar v_x | \psi_{ck} \rangle}{E_{ck} - E_{vk}}, \tag{3.3}
\]

and similarly for \(Y_{vcck}\). The second form in Eq. (3.3) makes use of the definition of the velocity operator as

\[
v = -\frac{i}{\hbar} [r, H], \tag{3.4}
\]

thereby taming the problematic position operator in the Bloch-state matrix element.

Averaging Eq. (3.2) over a unit cell of area \(A\), we find that the contribution of vertical coordinate \(z\) to the Chern number of the slab is

\[
C_z(z) = -\frac{4\pi}{A} \text{Im} \frac{1}{N_k} \sum_{vk} \sum_{vvc} X_{vcck} Y_{vc'ck}^\dagger \rho_{vc'k}(z), \tag{3.5}
\]

such that the total slab Chern number is \(C_z = \int dz C(z)\). In this equation,

\[
\rho_{vc'k}(z) = \langle \psi_{vk} | \left( \int_A d^2r |r\rangle \langle r| \right) |\psi_{vc'k}\rangle
\]

\[
= \int_A d^2r \psi_{vk}^*(r) \psi_{vc'k}(r) \tag{3.6}
\]

is the matrix element in the Bloch representation of the projector onto the slice through the unit cell at coordinate \(z\).

Because we work here in the tight-binding approximation, the sum over conduction states in Eq. (3.2) is a finite one and is easily carried out. We adopt a diagonal approximation to the matrix elements of the position operators in the tight-binding basis,

\[
\langle j|r|j' \rangle = \bar{r}_j \delta_{jj'} \text{ (here } j \text{ labels an orbital located at } \bar{r}_j \text{ in the unit cell of the slab)},
\]

in which case the numerator of Eq. (3.3) can be evaluated with the help of

\[
i\hbar \langle j|v_x|j' \rangle = (\bar{x}_j - \bar{x}_{j'}) H_{jj'} \tag{3.7}
\]

(and similarly for \(v_y\)), again making use of Eq. (3.4). Then the contribution of layer \(l\) to the Chern number of the slab follows the form of Eq. (3.5), becoming

\[
C_z(l) = -\frac{4\pi}{A} \text{Im} \frac{1}{N_k} \sum_{vk} \sum_{vvc} X_{vcck} Y_{vc'ck}^\dagger \rho_{vc'k}(l). \tag{3.8}
\]
Now
\[ \rho_{v\nu'k}(l) = \sum_{j \in l} \psi^*_{v'k}(j) \psi_{v'k}(j) \] (3.9)
is the Bloch representation of the projection onto layer \( l \), where the sum is over orbitals \( j \) belonging to that layer. All of the ingredients needed to compute Eq. (3.8) are thus easily evaluated, and the contribution of that layer to the AHC is just \( \sigma(l) = (e^2/h)C_z(l) \).

As explained in Rauch et al. [95], the object \( F_{v\nu'k} = (XY\dagger)_{v\nu'k} \) is the covariant metric-curvature tensor. Because \( \rho_k(l) \) is a Hermitian matrix, the imaginary part operation in Eq. (3.8) filters out only the antihermitian part of \( F_k \), which is \(-2i\) times the covariant Berry curvature tensor \( \Omega_{v\nu'k} \). Thus, formally we have that
\[ C_z(l) = \frac{2\pi}{A} \frac{1}{N_k} \sum_k \text{Tr}[\Omega_k \rho_k(l)] \] (3.10)
or, when converted to a Brillouin zone integral,
\[ C_z(l) = \frac{1}{2\pi} \int_{BZ} d^2k \text{Tr}[\Omega_k \rho_k(l)]. \] (3.11)

While this form is not as convenient for computational purposes as Eq. (3.8), it is more intuitive. For example, since \( \sum_l \rho_{v\nu'k} = \delta_{v\nu'} \), this immediately leads to \( C_z = \sum_l C_z(l) = \frac{1}{2\pi} \int_{BZ} \text{Tr}[\Omega_k] \) as it should.

In summary, we use Eq. (3.8) to compute the contribution \( C(l) \) of each surface layer to the surface AHC. If these vanish as \( z \) goes into the interior, then we simply sum the surface-layer contributions and multiply by \( e^2/h \) to get the surface AHC. If instead the \( C(l) \) oscillate in the bulk, then a coarse-graining procedure is used to isolate the surface contribution, as will be described later on.

### 3.2 Minimal model on the pyrochlore lattice

In this section we motivate and introduce a class of tight-binding models of spinors on the pyrochlore lattice.
3.2.1 The pyrochlore lattice: A playground for topological phenomena

In both the $A_2B_2O_7$ pyrochlore and $AB_2O_4$ spinel crystal structures, the $B$-site atoms form a simple inversion-symmetric pyrochlore lattice as shown in Fig. 3.1(a). Here we are interested in systems in which the metal atom on the $B$ is magnetic, while the one on the $A$ site is not. This means that our considerations will be limited to materials in which the $A$-site atom does not have a partially filled $f$ shell, or if it does, we are above the $f$-moment ordering temperature. In the case that a 5$d$ transition metal occupies the $B$ site, for example in pyrochlore $A_2Ir_2O_7$ iridates or spinel $AO_2O_4$ osmates, the 5$d$ electrons will exhibit strong spin-orbit coupling that can lead to band inversions. In fact, these 5$d$ electron systems have drawn a lot of attention recently, since the interplay between electronic correlation and that of spin-orbit coupling leads to a variety of topological and magnetic phases [24, 31, 32, 46, 67, 91, 104, 118, 130, 131, 142, 146, 147, 150, 153].

In the case of osmium spinels, it was shown [131] that for a reasonable range of the on-site Coulomb correlation $U$, the ground state can be a FM axion insulator. On the other hand, in the case of pyrochlore iridates it was shown [24, 130, 142] that electronic interactions can lead to topological phases with a noncollinear AIAO spin ordering, as shown in Fig. 3.1(b). Even though various experiments [31, 32, 104, 118] have confirmed that the ground state exhibits the AIAO phase, it was recently predicted [153] that pyrochlore iridates are topologically trivial in this phase.

3.2.2 The model

We can arrive at a simple tight-binding description of the pyrochlore iridates by considering the electronic states close to the Fermi energy. The oxygen octahedra surround the $Ir^{4+}$ ions, creating a strong cubic crystal field that splits the Ir 5$d$ orbitals into $t_{2g}$ and $e_g$ multiplets. The effective angular momentum of the $t_{2g}$ levels is $L_{\text{eff}} = 1$, so when the onsite SOC is “turned on,” it splits the $t_{2g}$ multiplet into an effective pseudospin $J_{\text{eff}} = 1/2$ doublet and a corresponding $J_{\text{eff}} = 3/2$ quadruplet [62], as illustrated in
Figure 3.1: (a) The pyrochlore lattice, comprised of an fcc lattice with a four-point basis (denoted by the numbers 1-4), forming a corner sharing tetrahedral network. (b) All-in-all-out configuration predicted for pyrochlore iridates. (c) The Ir 5$d$ level splitting in pyrochlore iridates due to the crystal field and the spin-orbit coupling. (d) Graphical definition of $\hat{b}_{ij}$ and $\hat{d}_{ij}$.

Fig. 3.1(c). Since Ir$^{4+}$ has five electrons in the 5$d$ shell, the $J_{\text{eff}} = 3/2$ quadruplet is filled, while the $J_{\text{eff}} = 1/2$ doublet is half-filled. Finally, the Fermi energy lies sufficiently far from the $J_{\text{eff}} = 3/2$ level so that, for the purposes of a minimal model, we can safely ignore it and describe the system in terms of a single spinor degree of freedom per site.

In the following we consider only nearest-neighbor interactions, with

$$H_t = t \sum_{\langle ij \rangle \sigma} c^\dagger_{i\sigma} c_{j\sigma} + \text{h.c.}$$

(3.12)

describing the spin-independent hoppings. For the spin-dependent interactions, we begin by imposing time-reversal symmetry, in which case the spin-orbit induced hopping between nearest neighbors takes the form

$$H_\lambda = \lambda \sum_{\langle ij \rangle \alpha \beta} i\sqrt{2} c^\dagger_{i\alpha} \hat{b}_{ij} \times \hat{d}_{ij} \cdot \sigma_{\alpha\beta} c_{j\beta} + \text{h.c.}$$

(3.13)
Here $\sigma = (\sigma_x, \sigma_y, \sigma_z)$ are the Pauli spin matrices, $\hat{d}_{ij}$ is the unit vector connecting site $i$ to site $j$, and $\hat{b}_{ij}$ is the unit vector from the center of a tetrahedron to the midpoint of the bond $\langle ij \rangle$ (see Fig. 3.1(d) for a graphical definition) [46, 67]. This structure is imposed by the presence of a mirror symmetry across the plane containing vectors $\hat{d}_{ij}$ and $\hat{b}_{ij}$, which ensures that only the component of $\sigma$ normal to this mirror plane can appear. The factor of $i$ in front is required by time-reversal symmetry, which reverses the signs of the Pauli matrices in addition to applying complex conjugation.

Next, in order to break time-reversal symmetry and model the magnetic ordering in these systems, we consider an onsite Zeeman term

$$H_\Delta = \Delta \sum_i \hat{n}_i \cdot \sigma c_i^\dagger c_i,$$

where $\hat{n}_i$, $i = \{1, 2, 3, 4\}$ are the vectors on each site describing the spin configuration.\(^1\) As mentioned earlier, the appearance of magnetic moments is a consequence of on-site Hubbard repulsion in the many-body Hamiltonian [130, 142], and the AIAO long-range order is established by the pattern of intersite exchange interactions [24]. Here we assume the presence of the AIAO magnetic order and regard our Hamiltonian in Eq. (3.14) as resulting from a mean-field approximation to the correlated system of interest.

### 3.3 Results

In the following, we first consider in detail the case of AIAO spin ordering, and then briefly describe how the results change when considering the FM ordering as a function of orientation of the magnetization. The calculation of bulk energy bands, surface states and Wannier bands was done using the built-in tools provided by the open-source code package PythTB. The PythTB scripts used to generate the results presented here are available in the Supplementary Material of Varnava and Vanderbilt [126]. In addition,

\(^{1}\text{Broken time-reversal symmetry also allows an } \hat{n}_{ij} \cdot \sigma c_i^\dagger c_j \text{ term to appear in the nearest-neighbor hopping, but the modification of the on-site Zeeman term is much more directly motivated.} \)
an extension package of PythTB containing a flexible routine for computing the layer-resolved AHC has been posted to an open-source repository.\footnote{PythTB scripts for computing layer-resolved contributions to the anomalous Hall conductivity are available at \url{http://www.physics.rutgers.edu/~dhv/pythtb-resources}.}

### 3.3.1 All-in-all-out spin configuration

**Figure 3.2:** (a) The first Brillouin zone with the path along which the bulk energy bands are calculated. (b)-(d) Energy bands for different values of the parameters \((t, \lambda, \Delta)\). (b) \((1, 0, 0)\). (c) \((1, 0.1, 0)\). (d) \((1, 0.1, 0.2)\).

**Bulk energy bands** — As we have seen in section 3.2.2, the Hamiltonian consists of three terms \(H(t, \lambda, \Delta) = H_t + H_\lambda + H_\Delta\). To get a better picture of the nature of the interactions, we turn our attention to the energy bands. First, we turn on only the spin-independent hoppings, \(H = H(t, 0, 0)\), obtaining the results shown in Fig. 3.2(b). Since spin up and down are equivalent, all bands are doubly degenerate, and the Hamiltonian is easily diagonalized to obtain the four solutions

\[
E_k^{(1,2)} = -2t \\
E_k^{(3,4)} = 2t[1 \pm \sqrt{1 + A_k}],
\]

where

\[
A_k = \cos 2k_x \cos 2k_y + \cos 2k_y \cos 2k_z + \cos 2k_z \cos 2k_x.
\]
(The extra degeneracy of the first two bands is a well-known artifact of this minimal model.) Since we are at half filling, this Hamiltonian describes a metal with a quadratic band touching at $\Gamma$.

Next we turn on the $\mathcal{T}$-invariant intersite spin-orbit coupling term, so that now $H = H(t, \lambda, 0)$. Even though the Hamiltonian contains spin-dependent interactions, the combination of $\mathcal{T}$ and $\mathcal{I}$ symmetries forces the bands to remain doubly degenerate everywhere. However, the other degeneracies are lifted except at some high-symmetry points, as shown in Fig. 3.2(c) for $\lambda = 0.1t$, where it is also clear that a global gap has opened between the valence and conduction bands. This insulating state cannot be connected adiabatically to the atomic limit and corresponds to a strong TI [46, 67] with $\theta_{cs} = \pi$.

Finally, we break $\mathcal{T}$ by turning on $\Delta$. For $\Delta$ small enough that the gap does not close, as in Fig. 3.2(d), $\theta_{cs}$ must remain equal to $\pi$. Now the topological phase is protected by inversion, and the system is an axion insulator. The breaking of $\mathcal{T}$ means that spin up and down electrons no longer disperse in the same way, and all eight bands are nondegenerate except at some symmetry points or lines in the Brillouin zone.

**Phase diagram —**

Using the Fu-Kane criterion discussed in the introduction, we will iterate over the parameters of our model to acquire the phase diagram shown in Fig. 3.3, where the numbers indicate the NOPS. Here we use the topological definition of an insulator, insisting only on a global direct gap between the four valence and four conduction bands. The trivial and non-trivial insulating states are separated by “Weyl Semimetal” regions that are represented by the red color.

The boundaries in Fig. 3.3 correspond to crossings of eigenvalues, i.e., band inversions, at the $\Gamma$ and $L$ points. All states at $\Gamma$ have the same (even) parity, and the parity counts at the $X$ points never change, so the NOPS is determined by the parity sum over the four $L$ points. Thus, the boundaries with a discontinuity in the NOPS involve band inversion at $L$, while the others are at $\Gamma$, and the Weyl semimetal regions lie in between.

Our picture, then, is as follows. Choose any path that starts from the axion phase,
Figure 3.3: Phase Diagram for $t = 1.0$ with the distribution of the total number of odd parity states at the TRIM. The Hamiltonian remains invariant under $\Delta \rightarrow -\Delta$ so it is sufficient to consider positive $\Delta$. On the other hand, $H(-t, -\lambda, \Delta) = -H(t, \lambda, \Delta)$ so for $t = -1.0$ the occupied and unoccupied bands are interchanged. The phase diagram is nevertheless the same, since we assume half filling and the 8 bands constitute a trivial insulator.

enters the Weyl semimetal phase, and ends in the trivial phase in Fig. 3.3. If you enter the Weyl region without changing the total NOPS, i.e., by passing through a Weyl semimetal region with 10 or 14 total NOPS, then you find that eight Weyl points are created at $\Gamma$ when crossing the boundary. These then separate and migrate along the eight equivalent $\Gamma \rightarrow L$ lines, and finally meet and annihilate in pairs at the four $L$ points at the crossing into the trivial insulator phase. If the path passes through a Weyl region with 12 total NOPS, on the other hand, the ordering of events is reversed.

Surface states — We create slabs that are in the axion-insulator phase and then use Eq. (3.8) to find the magnitude and sign of the surface AHC. Since, the bulk AHC is zero, and the surface AHC is not defined except for the component normal to the surface, we simply refer to it as $\sigma_{\text{AHC}}$. For the surface AHC to be half-integer, however, the surface band structure has to be gapped. In the axion phase there is no topological protection forcing surface states to cross the gap, but neither is an open gap guaranteed.
Figure 3.4: Surface band structures for (a-b) (111) surfaces, and (c-d) (001) surfaces. Blue (red) bands are localized on the bottom (top) surface. (a,c) Strong TIs, $H(t, \lambda, \Delta) = (1, 0.1, 0)$, with protected metallic surface states. (b,d) Axion insulators, $H(t, \lambda, \Delta) = (1, 0.1, 0.4)$; note gapped surface states on the (111) surface.

A slab along the (111) direction consists of alternating kagome and triangular atomic layers, so the 2D Brillouin zone is hexagonal. Fig. 3.4(a) shows the protected metallic surface band structure of a strong TI for a slab that terminates on a triangular layer. When the $\mathcal{T}$-breaking $\Delta$ term is turned on, the surface states become gapped as can be seen in Fig. 3.4(b).

Along (001) the slab consists of tetragonal layers rotated 90° with respect to each other, so the 2D Brillouin zone is tetragonal. As before, we choose the top and bottom layers to have the same orientation, so that the slab as a whole has inversion symmetry. In contrast with the (111) case, the surface states do not become gapped in the axion phase, as shown in Fig. 3.4(d). However, we have found that we can obtain an insulating surface by artificially increasing the strength of the Zeeman term on the surface atoms only, and this modified Hamiltonian will be used for some of the results presented below.
Surface anomalous Hall conductivity — We apply Eq. (3.8) for the slabs constructed along (111) and (001) directions and find the partial Chern numbers $C_z(l)$ of each layer, which we now denote just as $C(l)$ for brevity. The results are shown in Fig. 3.5(a) for slabs consisting of 20 layers and using a 40x40 k-mesh sampling. In both cases, the contribution to the AHC comes from the first few layers, and as one goes deeper in the bulk, the partial Chern number oscillates around zero. Note that the kagome layers contain three times as many atoms as the triangular layers Fig. 3.5(c), which may explain why the second layer along the (111) direction has approximately three times the partial Chern number of the first layer. In Fig. 3.5(b) we integrate the partial Chern number and show that $\sigma_{\text{AHC}}$ is quantized to half the quantum of conductance. To tame the oscillatory behavior of the bulk layers we carry out the sum to a depth $n$ according to

$$C_{\text{int}}(n) = \sum_{l=0}^{n-1} \tilde{C}_{l,l+1}$$

(3.17)

where $\tilde{C}_{l,l+1}$ is the coarse-grained average of two neighboring layers: $\tilde{C}_{l,l+1} = [C(l) + C(l + 1)]/2$ with the exception $\tilde{C}_{0,1} = C(1)/2$. In this way each layer is counted once except for the layer at depth $n$, which is counted with weight 1/2. This can be regarded as an application of the sliding window averaging method as described, e.g., in Appendix C of Resta [99]. Since these pairwise layer averages decay to zero in the bulk, the sum converges rapidly with depth $n$.

In both slabs we see that each time we add or remove a layer, the sign of $\sigma_{\text{AHC}}$ flips. For the (001) slab, we can understand this as follows. Adding a layer is equivalent to a 90° rotation around [001] and a half-lattice-constant translation along [001] followed by $T$, as is evident from Fig. 3.5(c). Neither the rotation nor the translation affects $\sigma_{\text{AHC}}$, but $T$ will flip its sign. Now for the (111) slab, things are less obvious since triangular and kagome terminations are inequivalent. Nevertheless, we can heuristically understand the sign flip by the expectation that the sign of the surface AHC tracks the sign of the surface magnetization, which reverses every time we add or remove a layer [see Fig. 3.5(c)].

Wannier bands —
Figure 3.5: (a) Layer-resolved partial Chern number $C(l)$ of Eq. (3.8) as a function of layer depth $l$ for slabs along (111) and (001) directions. For both slabs the bulk parameters are $(t, \lambda, \Delta) = (1.0, 0.1, 0.4)$ corresponding to the axion phase. In the case of the (001) direction the surface Zeeman term is modified to $\Delta_{\text{surf}} = 0.8$ in order to gap the surface states. (b) Integral of $C(l)$ over a surface region extending to depth $n$ as computed from Eq. (3.17). (c-d) Sketches of (001) and (111) slab orientations respectively for the AIAO spin configuration.

An analysis of the flow of the hybrid WCCs, otherwise known as Wilson loop eigenvalues, has been used extensively to study the topological properties of crystals [2, 5, 86, 111, 112]. For example, Taherinejad et al. [112] demonstrated how the Wannier bands of a strong TI must exhibit flow, no matter which direction is chosen for the Wannierization.

One may wonder, is the same is true of an axion insulator? Both axion insulators and strong TIs have the same non-trivial axion index, but axion insulators have broken $T$ symmetry. It turns out that the answer is no: a non-trivial axion index does not require flow.
Figure 3.6: WCCs for an axion insulator along the 4 projected TRIM for Wannierization directions (111) (a),(b) and (001) (c),(d). Figures (a),(c) correspond to a region in the phase diagram with 10 odd parity states at the TRIMs, where (b),(d) correspond to 14 odd parity states.

To demonstrate this, in Fig. 3.6 we plot the WCCs along paths connecting the four projected TRIM. We do this for two different Wannierization directions and two different parameter sets that describe an axion insulator. Figs. 3.6(b-c) show axion insulators that do not exhibit Wannier flow. Whether there is flow or not depends on how the number of odd parity states is distributed at the four projected TRIM. The nature of the Wannier bands in centrosymmetric crystals, and a discussion of how the axion $\mathbb{Z}_2$ invariant can often be deduced by an inspection of this Wannier structure, will be presented in a forthcoming publication.

3.3.2 Ferromagnetic spin configuration

It has been theoretically argued [131] that the ground state of osmium spinels, which share the same crystal structure as pyrochlores, is an FM axion insulator. From a topological point of view the FM and the AIAO spin configurations behave in a similar way, since both respect inversion symmetry. States that were in the strong TI phase for $\Delta = 0$ will become axion insulators when the ferromagnetism is turned on, up to a
critical Zeeman field of $\Delta_{c1}$ at which the bulk gap closes. Eventually, for $\Delta \gg t$ and $\lambda$, the system is a strongly spin-polarized insulator, with a gap separating spin-up from spin-down states, so that a second critical field $\Delta_{c2} > \Delta_{c1}$ must exist where the gap reopens. As in the AIAO case, we find that the surface states can be gapped in the FM axion-insulator phase, at least when the magnetization is normal to the surface, as shown in Fig. 3.7.

3.4 Surface Phenomena

In the previous sections we have constructed a model for an axion insulator and introduced a computational tool that enables us to calculate the surface AHC. In a sense, then, we have in hand a kind of virtual laboratory that we can use to explore the various phenomena that can occur on the surface of an axion insulator.

3.4.1 Surface AHC and magnetization direction

One interesting aspect of the FM configuration is that we can examine the behavior of $\sigma_{\text{AHC}}$ as we rotate the magnetization. For example, we know that the surface AHC wants to align with the magnetization, so when we rotate $\mathbf{M}$ from $[\bar{1}\bar{1}\bar{1}]$ to $[111]$ we should see a sign flip of $\sigma_{\text{AHC}}$. Of course as $\mathbf{M}$ is rotated, $\sigma_{\text{AHC}}$ has to vanish for
some critical angle $\vartheta_c$, and one would naively expect $\vartheta_c = \pi/2$. This is not the generic case, however; nothing forces $\sigma_{AHC}$ to vanish for $\vartheta = \pi/2$, as can be seen for example in Fig. 3.8(a), where we find that $\vartheta_c \approx 85^\circ$. Furthermore, when $\mathbf{M}$ is parallel to the surface, as in Fig. 3.8(b), $\sigma_{AHC}$ vanishes only for specific directions.

To understand Fig. 3.8(b), we turn our attention to the symmetries characterizing the slab. Neglecting $\mathcal{T}$ for the moment, we can obtain the “slab point group” by considering the subset of the bulk point group that maps the slab onto itself without interchanging the top and bottom surfaces, [112] i.e., preserving the (111) direction in the present case. Note that the pyrochlore point group contains three mirror planes containing the (111) axis, which therefore constitute good symmetries of the slab. To understand the role of $\mathbf{M}$, we turn our attention to the magnetic point group. $\mathbf{M}$ and $\sigma$ are odd under $\mathcal{T}$ and even under $\mathcal{I}$, so the mirror plane remains a good symmetry only if $\mathbf{M}$ is perpendicular to that mirror plane. In that case, a non-zero $\sigma_{AHC}$ is inconsistent with the symmetry, and therefore has to vanish. On the other hand, if $\mathbf{M}$ lies in one of the mirror planes, then the induced symmetry is mirror composed with $\mathcal{T}$ instead of a simple mirror, and this does not reverse $\sigma_{AHC}$. Thus, a non-zero $\sigma_{AHC}$ is allowed in this case, as well as for generic $\mathbf{M}$ directions for which all mirror symmetries are broken.

![Graph](image)

Figure 3.8: Surface AHC for a (111) slab as the magnetization is rotated: (a) in a plane perpendicular to the surface; (b) in the plane of the surface. The polar angle $\vartheta = 0$ corresponds to the $[\bar{1}\bar{1}\bar{1}]$ direction while the azimuthal angle $\phi = 0$ corresponds to the $[01\bar{1}]$ direction ($[01\bar{1}]$ is perpendicular to one of the mirror planes).
3.4.2 Termination-dependent surface AHC in the AIAO configuration

Figure 3.9: Illustration of possible surface AHC configurations for an AIAO insulator in the axion phase with (001)-type surface terminations. Blue and red colors represent positive and negative surface AHC respectively as defined in an outward-directed convention. (a) Removal of one atomic layer creates a chiral step channel. (b) Intersection of a domain wall and a chiral step channel results in a junction with two incoming and two outgoing chiral channels. (c) Intersection of facets with different signs of surface AHC generates chiral channels along the hinges. (d) Manipulation of surface terminations by addition or removal of layers can eliminate edge channels and lead to a configuration displaying the topological magnetoelectric effect (see text).

The surface AHC response of an axion insulator in the AIAO spin configuration is fascinating because it is at the same time robust, in the sense that it is half-integer, and sensitive, since adding or removing a layer flips the sign of the surface AHC.

Fig. 3.9(a) shows an axion insulator slab where a step has been created by removing a layer (or in general, an odd number of layers) over half the surface. The step is the boundary between two regions with opposite surface AHC, resulting in a chiral boundary mode with conductance $e^2/h$ located at this step. We adopt the convention of coloring the insulating surfaces red or blue according to the sign of the outward-directed
surface AHC in this and subsequent figures. This provides easy visual guidance to the location and direction of edge channels, which circulate clockwise around blue facets, i.e., those with positive surface AHC.

The same chiral channel could also be obtained at the intersection of the surface with an antiferromagnetic domain wall (AFM DW), as was first discussed by Mong et al. [79]. In Fig. 3.9(b) an AFM DW separates two regions with AIAO and all-out-all-in (AOAI) spin configurations. Since the AOAI configuration is related to the AIAO by $T$ symmetry, the resulting surfaces will have opposite AHC. An intersection of an AFM DW channel and a step channel results in a junction with two incoming and two outgoing channels, as in Fig. 3.9(b). A scattering matrix $T$ should describe how the outgoing amplitudes depend on the incoming ones, with conservation of charge guaranteeing that $T$ is unitary. An STM tip or other object in close proximity with the junction would cause a modification of the unitary scattering matrix, providing possibilities for the construction of a novel quantum switch or sensor.

Another interesting scenario comes into play when we think of macroscopic crystallites. At the edges (or “hinges”) where facets meet, the facets can have either the same or opposite signs of surface AHC. This is illustrated for the case of a cubic crystallite in Figs. 3.9(c-d). If one could somehow control the terminating layer at each of the six surfaces, then one could achieve the topological magnetoelectric effect by preparing all the surfaces with the same sign (i.e., color). A crystallite of volume $V$ would then exhibit the full quantized magnetoelectric polarizability of $e^2V/2\hbar$, and would be free of edge channels. In this case we have a rare example of a 3D bulk topological state with no bulk-boundary correspondence at all – on 2D surfaces, 1D edges, or 0D corners. In the language of “higher-order TIs,” this is like saying that the order is higher than the dimension of the bulk.

3.4.3 Higher-order topological classification of an FM axion insulators

As we have seen above, the half-integer surface AHC of an axion insulator may lead to chiral channels flowing along boundaries on its surface [79, 109]. For example, a macroscopic crystallite will have hinge states at the intersection of two facets with
opposite sign of the surface AHC, as was illustrated above. A similar discussion about higher-order states in an inversion-symmetric TI was recently presented in Khalaf [61]. Here we emphasize how these hinge-state configurations can be regarded as distinct topological phases, such that passing from one to the other, one has to pass through a metallic intermediate phase. In particular, some facets have to become metallic at such critical points. We will also illustrate how the surface phase diagram can be deduced for such a higher-order topological phase. For this purpose it is most convenient to focus on the case of the FM spin ordering, since this presents the opportunity for easy control via an externally applied magnetic field.

We start with an octahedral crystallite in the FM axion-insulator phase and imagine changing the direction of the magnetization. It is tempting to assume that as long as $M$ is not parallel to a facet, the perpendicular component of $M$ will open the surface state gap, with the sign of the half-integer surface AHC determined by the outward-normal component of $M$. By the same token, one may guess that if $M$ is parallel to a facet, $\sigma_{AHC} = 0$ and the facet will be metallic. While our results are broadly consistent with this picture, we have found that $\sigma_{AHC}$ does not necessarily pass through zero exactly at the expected 90° critical angle between $M$ and the surface normal, a fact that we shall return to shortly. For now, however, we adopt the broad picture in which the direction of $M$ determines the sign of the surface AHC on each facet in the expected way.

In what follows, we argue that the surface of an axion insulator exhibits multiple higher-order topological phases, and the surface phase diagram is described by a polyhedron embedded in the magnetization sphere, which serves as the parameter space. For example, the surface phase diagram of an octahedral crystallite in real space is described by a cuboctahedron in $M$ space, as shown in Figs. 3.10(a) and (b) respectively, while that of a cubic crystallite is described by an octahedron in $M$ space. In the former case, it is only when $M$ intersects an edge of this cuboctahedron that it becomes parallel to one of the crystallite facets, and the surface gap of that facet closes. On the other hand, when $M$ intersects a face of the cuboctahedron, there is a non-zero perpendicular component of $M$ on every crystallite facet, resulting in a particular coloring of all facets. When $M$ is as shown in Fig. 3.10(b), for example, the corresponding
Figure 3.10: Higher-order topological states at the surface of an FM axion insulator. (a) Octahedral crystallite with facets perpendicular to the (111) and equivalent directions. (b) Higher-order phase diagram. A given magnetization vector intersects the cuboctahedron at a particular face, with each face corresponding to a different arrangement of signs of the surface AHC on the crystallite. (c)-(e) Examples of surface configurations resulting from different magnetization directions. Color scheme follows that of Fig. 3.9.

This shows that each face of the cuboctahedron corresponds to a different topological phase of the crystallite, characterized by a particular chiral loop configuration flowing on the edges of the crystallite. The edges of the cuboctahedron then correspond to boundaries between two topological phases.

Realization of such a device would enable control of conducting channels using magnetic fields. For example, one can imagine attaching wires to the top, bottom, left, and right corners of the octahedron in Fig. 3.10(c-e). Rotating \( \mathbf{M} \) from \([100]\) to \([001]\) will then switch the system between configurations in which the horizontal or vertical wires are connected. Rotating \( \mathbf{M} \) to \([111]\) would instead connect all four wires. Here we have the potential for yet another kind of novel quantum switch.

We now return to a point mentioned earlier and recall that the picture presented
above needs a slight modification. That is, when tilting the magnetic field orientation, the closure of the surface gap may occur when the magnetic field is near, but not at, the condition of being parallel to the facet, unless this is enforced by some symmetry. An example was shown in Fig. 3.8(a), where the critical angle is around 85° instead of 90°, while Fig. 3.8(b) illustrates the role of symmetry. Taking this effect into account would cause the phase boundaries in Fig. 3.10(b) to become slightly distorted, with the edges of the cuboctahedron no longer being perfectly straight (i.e., no longer perfect great circles in M orientation space). Nevertheless, these distortions must respect the crystal symmetries, and have no effect on the qualitative aspects of the discussion given above.
Chapter 4

Axion coupling in the hybrid Wannier representation

For a $d$-dimensional insulator, the flow of the hybrid Wannier (HW) centers as a function of wavevector in the orthogonal $(d - 1)$-dimensional BZ, which we shall refer to as the Wannier band structure,\(^1\) often proves to be a very useful tool for determining its topology \([1, 2, 14, 55, 112, 126]\). For example, we have seen that the integer Chern number of a 2D QAH system, the $\mathbb{Z}_2$ index of a 2D TR-invariant insulator, and the strong and weak indices of a 3D TR-invariant insulator, are easily diagnosed via an inspection of the flow of the HW centers. They can also provide strong hints as to the location in the BZ of the band inversion responsible for the topological state, and to the flow of surface energy bands for surfaces orthogonal to the wannierization direction.

We have seen that inversion symmetry alone protects the axion $\mathbb{Z}_2$ index \([34, 53, 121]\). When TR is absent and the system is axion-odd, such systems are generally known as “axion insulators” \([130]\). A simple criterion was given by Turner et al. for determining the axion $\mathbb{Z}_2$ index in the presence of inversion symmetry \([121]\), generalizing the parity-counting analysis of Fu and Kane \([40]\) to the TR-broken case. Because the symmetry protecting the bulk topology (i.e., constraining the values of $\theta_{cs}$) is inversion instead of TR, and because inversion is never a good symmetry at any surface, it is much easier for the surface of an axion insulator to remain insulating. In some cases other symmetries may be present in addition and may force some facets to be metallic, and there is always the possibility that non-topological surface states will be present at the Fermi energy. Nevertheless, if the goal is to find insulators that naturally display a half-integer surface AHC response, then it appears that axion insulators are much more

---

\(^1\)For a 3D insulator, the bulk band structure is defined in the 3D BZ, while both the surface energy band structure and the Wannier band structure are defined in a projected 2D BZ.
promising. Unfortunately, physically realized examples of 3D bulk crystals that behave as axion insulators have been very difficult to find. To date the closest to a physical realization seems to the antiferromagnetic topological insulator \([79] \text{MnBi}_2\text{Te}_4\), which has been the subject of much recent interest \([23, 26, 71, 87, 149, 151]\).

Other symmetry operations also reverse the sign of \(\theta_{cs}\), and thus support an axion \(\mathbb{Z}_2\) classification. These include simple mirrors and glide mirrors, rotoinversions, and time-reversed rotations and screws. In general, an insulating material whose magnetic space group contains such axion-odd operators has an axion \(\mathbb{Z}_2\) index, and if that index is nontrivial, the material is guaranteed to have a half-integer QAH response on any insulating surface.

In this chapter, we explain how the axion coupling manifests in hybrid Wannier representation. We begin by describing the general properties of this manifestation, before we investigate the symmetry constraints imposed by the presence of axion-odd symmetry operations on the Wannier band structure, and more specifically, the additional constraints associated with the axion-odd topological state. Conversely, we show how the axion \(\mathbb{Z}_2\) index can often be deduced from an inspection of the Wannier band structure. In many cases, this involves only a visual inspection of the Wannier band structure, possibly including a counting of Dirac nodes between certain bands. In other cases it may require a calculation of the Chern number of a subset of Wannier bands, or a more complicated computation of Berry fluxes on truncated Wannier bands and Berry phases on the truncation boundaries. Finally we comment on the connection of this work with higher-order topology and bulk-edge correspondence.

Before we begin we want to establish some convention regarding the connectivity (sometimes called the “flow”) of the Wannier bands. We restrict ourselves to the case that all bulk Chern indices are zero. We refer to the vanishing of the Chern number in the \(x\)-\(y\) plane as the “in-plane Chern constraint.” The vanishing of the other two Chern indices guarantees that each Wannier band returns to itself (not to higher or lower partners) as one traverses the 2DBZ by a reciprocal lattice vector. Of course, it also returns to itself on any closed loop that does not wind by a reciprocal lattice.
Figure 4.1: Illustrative sketches of possible hybrid Wannier (i.e., Wilson loop) band structures for a model with four occupied bands. Hybrid Wannier (HW) centers $z_{nl}(k)$ repeat along the vertical direction with period $c$. The horizontal axis represents some path connecting high-symmetry points in the 2D Brillouin zone. (a) Four isolated Wannier bands. (b) Two connected groups of Wannier bands. (c) Fully connected Wannier bands. vector, since all such loops are contractible. Importantly, these considerations allow us to label the Wannier bands globally by integers that we take as increasing along $\hat{z}$.

A gap is said to exist between a pair of adjacent Wannier bands if these are not connected by degeneracies anywhere in the 2DBZ. A Wannier band is said to be isolated if a gap exists above and below it. A connected group of Wannier bands is a set of adjacent bands that are connected by degeneracies, but that are separated by gaps above and below the group. The HW sheet structure as a whole is said to be connected if there are no gaps; otherwise it is disconnected and is composed of $M$ internally connected groups separated by $M$ gaps per unit cell along $z$. These features are illustrated in Fig. 4.1.

4.1 General properties

In this section we develop rather general expressions for the axion coupling in the hybrid Wannier representation, valid even in the case of degeneracies between Wannier bands. At this stage we make no special assumptions about symmetries, so the axion coupling

---

2The fact that the bulk is globally insulating is also important. In a Weyl semimetal, for example, traversing a loop that winds around the projection of a Weyl point in the 2DBZ pumps the Berry phase by $2\pi$, so the Wannier bands cannot return to themselves around such a loop.
is not necessarily quantized.

4.1.1 Disconnected Wannier band structure

Isolated bands — Let us begin with the simplest case, in which all Wannier bands are isolated, so that \( z_{ln} \) is a smooth function of \( \mathbf{k} \) for each \( n \). For this case, which is illustrated in Fig. 4.1(a), the authors of Taherinejad and Vanderbilt [111] and Olsen et al. [86] showed that the CS axion coupling \( \theta_{cs} \) can be expressed in the HW representation as

\[
\theta_{cs} = \theta_z \Omega + \theta_{\Delta xy}.
\]  

(4.1)

The first term

\[
\theta_z \Omega = -\frac{1}{e} \int d^2k \sum_n z_{0n} \Omega_{0n,0n}
\]  

(4.2)

is a kind of “Berry curvature dipole” term, where \( z_{0n} \) is given by Eq. (2.19) and the summation index \( n \) runs over Wannier bands in the home unit cell (\( l=0 \)). The second term in Eq. (4.1) can be written as

\[
\theta_{\Delta xy} = -\frac{1}{e} \int d^2k \sum_n \sum_{l'n'} \Gamma_{0n,l'n'}
\]  

(4.3)

where

\[
\Gamma_{ln,l'n'} = i (z_{l'n'} - z_{ln}) A_{ln,l'n'}^x A_{l'n',ln}^y
\]  

(4.4)

which is fully gauge invariant, based on the transformation of the Berry connection of the hybrid Wannier states [224] i.e.,

\[
\tilde{\Gamma}_{ln,l'n'} = \Gamma_{ln,l'n'}.
\]  

(4.5)

To see this, note that the \( (z_{l'n'} - z_{ln}) \) prefactor in Eq. (4.4) insures that only off-diagonal elements of \( A^\alpha \) contribute, and that the product of \( A_{ln,l'n'}^x \) and \( A_{l'n',ln}^y \) leads to a cancellation of the phase factors in Eq. (2.24).

We have chosen to fix index \( l=0 \), but the result is unchanged if the index 0 is replaced by arbitrary \( l \) in Eq. (4.3), since \( \Gamma \) obeys the same kind of translational invariance as in Eq. (2.22), i.e.,

\[
\Gamma_{ln,l'n'} = \Gamma_{0n,(l'-l)n'}.
\]  

(4.6)
Henceforth we simplify the notation by establishing the convention that the absence of a cell index \( l \) implies \( l = 0 \), i.e., \( z_n \) is shorthand for \( z_{0n} \), \( \Omega_{nn} \) is shorthand for \( \Omega_{0n,0n} \), etc. Then Eq. (4.2) becomes just

\[
\theta_{z\Omega} = -\frac{1}{c} \int d^2k \sum_n z_n \Omega_{nn}.
\] (4.7)

Similarly, \(|h_n\rangle\) shall refer to a HW function \(|h_{0n}\rangle\) in the home unit cell \( l = 0 \).

While all the quantities in Eqs. (4.3) and (4.7) are gauge invariant, so that there is no ambiguity on this account, there is an important ambiguity of a different kind in \( \theta_{z\Omega} \). It comes from the freedom to choose the set of Wannier bands assigned to the home unit cell; this affects \( \theta_{z\Omega} \) because of the appearance of \( z_n \) in Eq. (4.7). By contrast, only differences of \( z \) values appear in Eq. (4.3), so this term is fully unambiguous.

To understand the ambiguity in \( \theta_{z\Omega} \), recall that the labels \( n = \{1, 2, \ldots, J\} \) simply count the bands in ascending order within the home unit cell. (As a reminder, we are considering here the case that all Wannier bands are isolated.) But how do we choose the unit cell? Which band shall we label as \( n = 1 \)? One way to think about this is that the choice of home unit cell corresponds to the choice of one of the \( J \) gaps as the “primary” one, such that the counting starts with the Wannier band just above this gap. A different choice of primary gap just has the effect of shifting some subset of the Wannier bands along \( z \) by distance \( c \). For each shifted band \( n \), \( \theta_{z\Omega} \) in Eq. (4.7) gets shifted by

\[
\Delta \theta_{z\Omega} = -\int d^2k \Omega_{nn} = -2\pi C_n,
\] (4.8)

where \( C_n \) is the Chern number computed over this Wannier band. Since \( C_n \) is necessarily an integer, this means that \( \theta_{z\Omega} \), and also the total \( \theta_{cs} \) as given by Eq. (4.1), is only well defined modulo \( 2\pi \). Actually, such an ambiguity is expected, since we know on other grounds that \( \theta_{cs} \) is only well defined modulo \( 2\pi \), so this is not problematic. However, it is still something that we have to anticipate and deal with in the analysis that follows.

To clarify why \( \theta_{\Delta xy} \) is unaffected by the choice of unit cell, it is instructive to rewrite Eq. (4.3) as

\[
\theta_{\Delta xy} = -\frac{1}{Nc} \int d^2k \sum_{\mu\mu'} \Gamma_{\mu\mu'},
\] (4.9)
where we have introduced a condensed index notation $\mu = (ln)$, the sums over $\mu$ and $\mu'$ both run over all Wannier bands in a large system of $N$ unit cells, and $\Gamma_{\mu\mu'} = i(z_{\mu'} - z_{\mu})A^x_{\mu}A^y_{\mu'}$. In other words, $\theta_{\Delta xy}$ is just the sum of all $\Gamma$ elements taken per unit cell along $\hat{z}$. With this perspective, it is obvious that the $ln$ labeling of the Wannier bands is irrelevant for $\theta_{\Delta xy}$. By contrast, it is impossible to write an expression similar to Eq. (4.9) for $\theta_{z\Omega}$, since the appearance of $z_{\mu}$ itself, rather than the difference $z_{\mu'} - z_{\mu}$, would render the average over $N$ cells ill-defined.

*Composite groups of bands* — We now consider the case that at least some of the bands form internally connected composite groups, but the Wannier band structure as a whole remains disconnected. An example of a system of this type is shown in Fig. 4.1(b), where there are $M = 2$ connected groups, each consisting of a pair of bands joined by a nodal point. We do not expect such nodal points to appear generically, since accidental degeneracies have codimension three, and thus require fine tuning. On the other hand, such nodal degeneracies may sometimes be induced by symmetry, often occurring at the time-reversal invariant momenta (TRIM), or at other high-symmetry points or lines, in the 2DBZ. When such degeneracies are present, the evaluation of Eq. (4.1) becomes problematic because $\Omega_{nn}$ can diverge in the vicinity of the degeneracies between Wannier bands. In the case of a Dirac node, defined as an isolated nodal touching with linear dispersion of $z_n(k)$ close to the node, $\Omega_{nn}$ has a delta-function singularity at the node. To solve this problem, we can go over to a formulation in terms of a gauge-covariant treatment of the Berry curvature within each group.

To see this, let the $a$'th group ($a = 1, \ldots, M$) be composed of $M_a$ Wannier bands, and combine terms such that the contribution of group $a$ is taken to be

$$\theta_a = -\frac{1}{c} \int d^2k \left( \sum_{n \in a} z_n \Omega_{nn} + \sum_{n, n' \in a} \Gamma_{nn'} \right).$$

As a reminder, $z_n, \Omega_{nn'},$ and $\Gamma_{nn'}$ refer to contributions coming from the home cell $l = 0$ (and $l' = 0$) only. Then the total Chern-Simons coupling is

$$\theta_{cs} = \theta'_{z\Omega} + \theta'_{\Delta xy}$$

(4.11)
where

$$\theta'_z \Omega = \sum_a \theta_a$$

(4.12)

is the sum of Eq. (4.10) over groups $a$, and

$$\theta'_{\Delta xy} = -\frac{1}{c} \int d^2k \sum_n \sum'_{n'} \Gamma_{0n,0n'}$$

(4.13)

is identical to $\theta_{\Delta xy}$ in Eq. (4.3) except that the prime on the sum indicates the omission of all terms with Wannier bands $l'n'$ belonging to the same group as $0n$. The sum over groups in Eq. (4.12) counts isolated bands by treating them as groups with $M_a = 1$, and Eq. (4.1) with Eqs. (4.3) and (4.7) are recovered if all bands are isolated.

But now the quantity inside the parentheses in Eq. (4.10) can be simplified by writing it as

$$\theta_a = -\frac{1}{c} \int d^2k \sum_{n \in a} z_n \tilde{\Omega}_{nn}$$

(4.14)

where

$$\tilde{\Omega}_{nn} = \Omega_{nn} - i \sum_{m \in a} (A^{x}_{nm} A^{y}_{mn} - A^{y}_{nm} A^{x}_{mn})$$

(4.15)

is the diagonal element of a gauge-covariant Berry curvature matrix. Substituting Eq. (4.15) into Eq. (4.14) easily demonstrates the equivalence of these expressions. The Chern number of a connected group of bands can then be expressed in terms of the gauge-covariant Berry curvature as

$$C_a = \frac{1}{2\pi} \int d^2k \sum_{n \in a} \tilde{\Omega}_{nn}.$$  

(4.16)

The advantage of this formulation is that $\tilde{\Omega}_{nn}$ remains a smooth and divergence-free function of $k$ in the vicinity of degeneracies between bands in the group. This is a well-known feature of the gauge-covariant Berry curvature, and can be seen by expressing it as

$$\tilde{\Omega}_{nn} = -2 \text{Im} \langle \partial_x h_n | \partial_y h_n \rangle,$$

(4.17)

where $\partial_x |h_n\rangle = Q_a \partial_x |h_n\rangle$ is the gauge-covariant derivative of the HW state and $Q_a = 1 - \sum_{n \in a} |h_n\rangle \langle h_n|$ is the projector onto all states other than the Wannier bands in group $a$ in the home unit cell. (The philosophy is similar to that used in defining the gauge-covariant derivative $\partial_{\alpha} |u_{nk}\rangle = Q_{nk} \partial_{\alpha} |u_{nk}\rangle$ of the Bloch energy eigenstate $|u_{nk}\rangle$. )
where $Q_{n\mathbf{k}} = 1 - \sum_{m\neq n} |u_{m\mathbf{k}}\rangle \langle u_{m\mathbf{k}}|$, except that here we work with the spectrum of $z$, not that of $H$. The projector $Q_a$ eliminates from $\tilde{\partial}_a |h_n\rangle$ the divergences that would be present in $\partial_a |h_n\rangle$ arising from mixing between Wannier bands inside the same group.

In summary, in this part we have argued that Eqs. (4.11-4.14) form a robust set of equations that can be used to evaluate the CS coupling even in the presence of composite groups of internally-degenerate Wannier bands.

In case there is a doubt about the correctness of this expression, we can consider the application of a small symmetry-lowering perturbation $\lambda V$ that gaps out the degeneracies between bands within the group. In this case we know Eq. (4.1) is correct, and we can argue that it is equivalent to Eq. (4.11), and then take the limit $\lambda \rightarrow 0$. Since Eq. (4.11) is insensitive to degeneracies within the group, we can conclude that its evaluation at $\lambda = 0$ provides the correct CS coupling.

### 4.1.2 Introduction of a cutting surface

![Figure 4.2](image)

Figure 4.2: (a) An isolated Wannier band (green) lying entirely inside the home unit cell (cutting surfaces, blue, at $z = 0$ and $z = c$). (b) The same Wannier band, now represented by two disconnected pieces in the home unit cell, after the cutting surface has been raised. (c) The region $S_n$ where the Wannier band falls below the new cutting surface, and its boundary $C_n$.

When it comes time to consider the case of a fully connected Wannier band structure, it will not be possible to assign Wannier bands to the home unit cell without cutting through the Wannier bands in some way. To prepare for this, we begin by returning to the case of isolated Wannier bands, and we define a “cutting surface” $z_{cut}(\mathbf{k})$ that
is smooth and periodic in $k$, such that all Wannier bands with $z_{\text{cut}}(k) < z_n(k) < z_{\text{cut}}(k) + c$ are assigned to the home cell $l = 0$. As a reference, if the cutting surface lies entirely inside the primary gap, as illustrated in Fig. 4.2(a), then the definition of the home unit cell is the same as it was in Sec. 4.1.1.

Instead, let $z_{\text{cut}}(k)$ be increased such that it cuts through one or more of the Wannier bands, as illustrated for a single Wannier band in Fig. 4.2(b). This has the effect that $z_n$ is shifted upwards by $c$ for all bands lying below the new cut. Following an earlier argument, the contribution of Wannier bands lying entirely below the cut is changed by $2\pi$ times a Chern integer, making no change to $\theta_{cs}$ modulo $2\pi$. As for band $n$ pierced by the cut, we define $S_n$ to be the region of the 2D plane for which $z_n$ lies below the cut, and $C_n$ is its boundary, i.e., the intersection with the cut, as shown in Fig. 4.2(c). The reassignment of the HW states inside $S_n$ shifts $z_n$ upwards by $c$, so that the term $\theta_{z\Omega}$ in Eq. (4.7) changes by an amount

$$\Delta \theta_{z\Omega}^{(n)} = - \int_{S_n} d^2k \Omega_{nn} = -\phi_{\text{cut}}^{(n)}$$

where

$$\phi_{\text{cut}}^{(n)} = \oint_{C_n} A_{nn} \cdot dk$$

(4.19)

is the Berry phase evaluated on $C_n$. In general, the boundary $C_n$ could be multiply connected, in which case the sum over loop Berry phases is implied in Eq. (4.18). The total change in $\theta_{z\Omega}$ is then

$$\Delta \theta_{z\Omega} = -\phi_{\text{cut}} = - \sum_n \phi_{\text{cut}}^{(n)} ,$$

(4.20)

where $\phi_{\text{cut}}$ is the total Berry phase from all Wannier bands that intersect $z_{\text{cut}}(k)$. Of course, this quantity is only well defined modulo $2\pi$, but this is not an issue since the axion coupling has the same indeterminacy.

On the other hand, the expression for $\theta_{\Delta xy}$ in Eq. (4.3) is unchanged, since no matter the labeling, all pairs of Wannier bands at the same $k$ eventually enter the sum in Eq. (4.3) in the same way as before. The overall change in Eq. (4.1) is then just $\Delta \theta_{cs} = -\phi_{\text{cut}}$. To correct for this, we just have to add back a piece to cancel Eq. (4.20), and we arrive at

$$\theta_{cs} = \theta_{z\Omega} + \theta_{\Delta xy} + \phi_{\text{cut}} .$$

(4.21)
As a reminder, $\theta_{z\Omega}$ is still evaluated as in Eq. (4.7), but now with the band label $n$ running from 1 to $J$ independently at each $k$ beginning with the first band above the cutting surface, and the Berry-phase term $\phi_{\text{cut}}$ accounts for the contributions of the loops of intersection of the cutting surface with the bands. Equation (4.21) is one of the principal results of the present work.

### 4.1.3 Connected Wannier band structure

The result in Eq. (4.21) was derived for the case that all bands are isolated, but we now wish to consider a fully connected Wannier band structure, in which no bands are isolated and no gaps occur.

**Degeneracy regions** — As a first step, we revise Eq. (4.21) for the case of composite groups. In particular, we consider the case that the cutting surface $z_{\text{cut}}(k)$ cuts through one or more of the connected bands within a group (but avoiding degeneracies). Unfortunately, we cannot simply cut through a connected group that is being treated using the gauge-covariant formulation considered for composite groups, because $\phi_{\text{cut}}$ is defined as the Berry phase on an individual Wannier band; this is equal to the integral of the $\Omega_{nn}$ over the enclosed area, but not that of $\tilde{\Omega}_{nn}$.

![Figure 4.3](image)

Figure 4.3: Sketch of degeneracy region (shaded) associated with two Wannier bands $z_n(k)$ near a point node, and its projection onto the disk-shaped region $\mathcal{R}_\alpha$ in the 2DBZ (bottom).
To remedy this problem, we can adopt a more restrictive treatment of degeneracies, as follows. Wherever there is a degeneracy between Wannier bands, we identify a degeneracy region (DR) surrounding the degeneracy. The $\alpha$’th DR consists of a set of $\mathcal{M}_\alpha$ adjacent bands ($\mathcal{M}_\alpha \geq 2$) that are involved in the degeneracy, and a small region $\mathcal{R}_\alpha$ surrounding the degeneracy in the 2DBZ, as illustrated in Fig. 4.3. If a cutting surface $z_{\text{cut}}(k)$ is present, we shall insist that it be chosen to avoid all the DRs, so that each DR lies entirely inside the home unit cell. We then collect together the terms in Eqs. (4.1-4.3) that only involve bands inside the DR to get a contribution

$$\theta^{\text{DR}}_\alpha = -\frac{1}{c} \int_{\mathcal{R}_\alpha} d^2k \left( \sum_{n \in \alpha} z_n \Omega_{nn} + \sum_{n,n' \in \alpha} \Gamma_{nn'} \right)$$

$$= -\frac{1}{c} \int_{\mathcal{R}_\alpha} d^2k \sum_{n \in \alpha} z_n \tilde{\Omega}_{nn} . \quad (4.22)$$

Here we have followed the same approach leading to Eq. (4.10), but now restricting the $k$ integral only to the region $\mathcal{R}_\alpha$, and the sum to run only over the $\mathcal{M}_\alpha$ bands involved in the degeneracy.

We then write the total Chern-Simons axion coupling in Eq. (4.21) as

$$\theta_{\text{cs}} = \theta''_{z\Omega} + \theta''_{\Delta xy} + \phi_{\text{cut}} , \quad (4.23)$$

where

$$\theta''_{z\Omega} = -\frac{1}{c} \int d^2k \sum''_n z_n \Omega_{nn} + \sum_\alpha \theta^{\text{DR}}_\alpha \quad (4.24)$$

and

$$\theta''_{\Delta xy} = -\frac{1}{c} \int d^2k \sum_{n} \sum''_{l,l'} \Gamma_{0n,0ll'} . \quad (4.25)$$

The double prime on the first sum in Eq. (4.24) indicates that all terms coming from within a DR are to be omitted, and in Eq. (4.25) it excludes terms where both bands lie in the same DR in the same cell; these omissions are compensated by the second term in Eq. (4.24).

Equation (4.23) is the desired formula, which remains robust in the presence of degeneracies, including in the connected case in which no gaps are present. It provides the formal solution to the problem of expressing the axion coupling $\theta_{\text{cs}}$ in the HW
representation, even in the case of fully connected Wannier bands, and is one of our principal results.

While Eq. (4.23) might be somewhat awkward to implement in practice, involving as it does the choice of some DRs that need to be treated differently while integrating over the 2DBZ and summing over bands, we shall mainly be interested below in cases in which some symmetry is present that quantizes $\theta_{cs}$ to 0 or $\pi$. In such cases, we shall insist on choosing the cutting surface and the DRs in such a way as to respect those symmetries, so that the same kinds of symmetry arguments used for the HW sheet contributions can also be used for the DR contributions. Thus, in cases where the $\theta_{z\Omega}$ contribution would vanish in the absence of DRs, $\theta''_{z\Omega}$ vanishes in their presence as well. In other cases, we shall argue shortly that one can take a limit in which the size of the DRs goes to zero. Either way, an explicit calculation of the contribution of a DR can typically be avoided. In fact, it often happens that only the last term $\phi_{cut}$ survives in Eq. (4.23), so that the axion coupling is given just by the Berry phase on the cutting loop (or the total Berry phase in the case of multiple loops).

Of course, whenever the Wannier band structure is actually disconnected, it is simpler to return to Eq. (4.11), where no cutting surface is needed and the bands can be indexed by counting from above some chosen gap. Again, symmetry will often allow us to decide the value of $\theta_{cs}$ based on rather general features of the Wannier band structure in this case as well.

**Shrinking the degeneracy regions** — Up to this point, we have avoided specifying the nature of the degeneracies between Wannier bands, which in general could occur at point nodes, or along lines, or even over a plane spanning the 2DBZ, depending on the type of symmetries present. Henceforth we will focus on point touchings of two or more Wannier bands, commenting only occasionally on the case of higher-dimensional degeneracies. Then, from a formal point of view, we can shrink the size of the $\alpha$'th DR surrounding a point node to a disk of some small radius $\epsilon$ in the 2DBZ. This may be problematic computationally, since the immediate vicinity of the DR may become difficult to treat without the gauge-covariant formulation, but as a formal manipulation it is permissible. Furthermore, we can argue that the contribution $\theta^{DR}_{\alpha}$ vanishes in the...
limit that $\epsilon \to 0$, since $\tilde{\Omega}_{nm}$ remains finite as the degeneracy is approached (since it only “sees” Wannier bands outside the degenerate group), and the area of the disk goes to zero.

Formally speaking, then, we can simply neglect the contributions from the DRs in the small-DR limit. This will prove useful in analyzing the contributions to $\theta_{cs}$ in the presence of certain symmetries, as we shall see.

\section{Symmetry considerations}

For the remainder of the manuscript, we restrict ourselves to insulating systems with axion-odd symmetries, which are classified by the axion $\mathbb{Z}_2$ index. In this section we apply the formalism developed above to such cases. We consider three different classes of axion-odd symmetries, explaining when and how the axion $\mathbb{Z}_2$ index can be inferred from an inspection of the Wannier band structure. When it cannot, we indicate what is the additional information that is needed to determine the axion $\mathbb{Z}_2$ index in the HW representation.

\subsection{Axion-odd insulators}

For the remainder of the manuscript we restrict ourselves to insulating systems with symmetries that protect the quantization of $\theta_{cs}$ to 0 or $\pi$. This symmetry could be time reversal (TR), in which case a spinful system with $\theta_{cs} = \pi$ is usually denoted as a strong topological insulator (TI). It also could be inversion $I$, in which case the system is usually called an axion insulator. However, many other symmetries can quantize $\theta_{cs}$, such that $\theta_{cs}/\pi = 0$ or 1 defines an “axion $\mathbb{Z}_2$ index.” We use the term “axion-odd insulators” to refer to systems in which the nontrivial $\mathbb{Z}_2$ index is protected by one of these symmetries, with TR-protected strong TIs and inversion-protected axion insulators as special cases.

In this section we survey the symmetries that can quantize the axion coupling, and describe their consequences for the Wannier band structure. Moreover, we show that in many cases it is possible to determine the axion $\mathbb{Z}_2$ index without recourse to the
integral expression for $\theta_{cs}$. In particular, it is often enough just to have a knowledge of some elementary features of the Wannier bands, such as the type and number of touchings between bands, or the total Chern numbers of certain bands or band groups.

To decide whether $\theta_{cs}$ is quantized to 0 or $\pi$ by a set of crystal symmetries – i.e., whether the axion $Z_2$ index is protected – we can just look at whether there are any elements in the magnetic point group that reverse the sign of $\theta_{cs}$. We shall call these the “axion-odd” symmetries, and they are comprised of the proper rotations composed with TR and the improper rotations not composed with TR. If one or more of these symmetries is present in the magnetic point group, then $\theta_{cs}$ is quantized to be 0 or $\pi$, i.e., the $Z_2$ index exists.

In such a case, we can argue as follows that the $\theta_{\Delta xy}$ term in Eq. (4.3) must vanish. Like the $\theta_{z\Omega}$ term, $\theta_{\Delta xy}$ has its sign reversed by any axion-odd symmetry operation. On the other hand, unlike $\theta_{z\Omega}$, $\theta_{\Delta xy}$ has no quantum of ambiguity, as we saw for the case of isolated bands. As a result, $\theta_{\Delta xy}$ is immediately forced to vanish in the presence of such a symmetry, while $\theta_{z\Omega}$ is not. The restricted sums $\theta'_{\Delta xy}$ and $\theta''_{\Delta xy}$ in Eq. (4.11) and (4.23), respectively, will vanish as well, since the assignment of bands to composite groups automatically respects the symmetry, and we chose the shapes and locations of the degeneracy regions to respect the symmetry as well. For this reason, we ignore the $\theta_{\Delta xy}$ terms in the considerations that follow, and the remaining question is whether the $\theta_{z\Omega}$ term, taken together with $\phi_{\text{cut}}$ if a cutting surface is present, yields 0 or $\pi$.

In many cases the magnetic point group may contain several axion-odd symmetry operations, but any one of them is enough to quantize $\theta_{cs}$. Therefore, we shall just consider each kind of axion-odd point symmetry in turn, and study its consequences, while keeping in mind that other symmetries may exist as well.

Let $g$ be the axion-odd magnetic point symmetry in question. For this $g$, we choose our Cartesian frame such that the direction $\hat{z}$ is either invariant or reversed by $g$, and then we carry out the wannierization along the $\hat{z}$ direction. In some cases this choice can be made in more than one way; for example, for a mirror, $\hat{z}$ could be chosen to lie in, or normal to, the mirror plane. In such cases more than one avenue of investigation may be available, with one possibly being more advantageous, but for now we just
Table 4.1: Axion-odd point-group symmetries, and their decomposition as $g = g∥g⊥$.

<table>
<thead>
<tr>
<th>$g$</th>
<th>$g∥$</th>
<th>$g⊥$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M_z$</td>
<td>$M_z$</td>
<td>$E$</td>
</tr>
<tr>
<td>$I$</td>
<td>$M_z$</td>
<td>$C_2$</td>
</tr>
<tr>
<td>$S_{3,4,6}$</td>
<td>$M_z$</td>
<td>$C_{3,4,6}$</td>
</tr>
<tr>
<td>$\bar{C}_2'$</td>
<td>$M_z$</td>
<td>$M_d'$</td>
</tr>
<tr>
<td>$E'$</td>
<td>$E$</td>
<td>$E'$</td>
</tr>
<tr>
<td>$C_{2,3,4,6}'$</td>
<td>$E$</td>
<td>$C_{2,3,4,6}'$</td>
</tr>
<tr>
<td>$M_d$</td>
<td>$E$</td>
<td>$M_d$</td>
</tr>
</tbody>
</table>

Operations reversing $\hat{z}$

Operations preserving $\hat{z}$

assume that some such choice has been made.

In this chosen frame, the possibilities for the spatial part of the point-group operator can be enumerated as follows. If $\hat{z}$ is not reversed, it is the identity $E$; a proper rotation $C_n$ by $2\pi/n$ about the $\hat{z}$ axis; or a reflection $M_d$ across a plane containing the $\hat{z}$ axis. If $\hat{z}$ is reversed, it is the inversion $I$; a mirror $M_z$ across the $x$-$y$ plane; an improper rotation $S_n = M_zC_n$ for $n = \{3, 4, 6\}$; or a two-fold rotation $\bar{C}_2$ about an axis lying in the $x$-$y$ plane. We attach a prime to denote composition with TR, so that $E'$ represents TR itself. This establishes our notations for the symmetry operations. We also decompose $g = g∥g⊥$, where $g∥$ is $E$ or $M_z$, and $g⊥$ is the in-plane part of the operator, including the TR component if present. For orientation, we list in Table 4.1 the axion-odd point-group symmetry operations $g$ along with their decomposition into $g∥$ and $g⊥$.

Keeping in mind that we are restricting ourselves to the case that $g$ is axion-odd, we can distinguish three general cases:

1. Operation $g$ reverses $\hat{z}$. We assume a choice of $z$-origin such that the corresponding space-group operation has no fractional translation along $\hat{z}$.³

³If the original choice of origin is such that the operation takes $z \rightarrow z_0 - z$, then we reset the origin
2. Operation \( g \) preserves \( \hat{z} \), and the corresponding space-group operation has no fractional translation along \( \hat{z} \).

3. Operation \( g \) preserves \( \hat{z} \), and the corresponding space-group operation involves a fractional translation \( c\hat{z}/m \).

The first case corresponds to \( g_{||} = M_z \), while the second and third pertain to \( g_{||} = E \). In all cases there may also be a fractional (nonsymmorphic) in-plane translation \( \tau_\perp \) in the space-group operation associated with \( g \); these play a role in determining when and where point nodes of degeneracy may occur between neighboring Wannier bands, but will not concern us here. If point nodes do occur, these will typically be located at \( z = 0 \) or \( z = c/2 \) for \( g_{||} = M_z \), and at general locations for \( g_{||} = E \).

We denote the axion-odd space-group operations as being either \( z \)-reversing, \( z \)-preserving, or \( z \)-nonsymmorphic according to the cases enumerated above. In the following subsections we shall consider each of these cases in turn.

Before proceeding, we list the transformation rules for the Wannier bands and their Berry curvatures. Under an operation \( \{g|c\hat{z}/m + \tau_\perp\} \) of the space group, a Wannier band transforms as [112]

\[
z_{n'}(\pm g_{\perp}k) = g_{||}z_n(k) + c\hat{z}/m, \tag{4.26}
\]

where the minus sign applies when \( g_{\perp} \) includes TR. The Berry curvature is a pseudovector pointing along \( \hat{z} \), and as a result it transforms as

\[
\Omega_{n'n'}(\pm g_{\perp}k) = \begin{cases} 
\Omega_{nn}(k), & g \text{ reverses } \hat{z}, \\
-\Omega_{nn}(k), & g \text{ preserves } \hat{z},
\end{cases} \tag{4.27}
\]

and the same rules apply to the gauge-covariant Berry curvature \( \tilde{\Omega}_{nn} \).

### 4.2.2 Determination of the quantized axion coupling

**Symmetry operation reverses \( \hat{z} \) —** For the operations that reverse \( \hat{z} \), listed in the first four rows of Table 4.1, it follows from Eq. (4.26) that each HW center at \( (z, k) \) in band at the invariant point \( z_0/2 \).
n has a partner at \((-z, \pm g_\perp k\)) in some band \(n'\) (again, the minus sign applies when \(g_\perp\) includes TR). Moreover, according to Eq. (4.27) the Berry curvatures of the Wannier bands are identical at these locations. At least in a simple situation such as that in the case of isolated bands, it therefore appears at first sight that all contributions to the \(\theta_{z\Omega}\) term given by Eq. (4.7) will cancel in pairs when summing over all Wannier bands and integrating over the 2DBZ, because of the sign reversal of \(z\). However, it is not always possible to choose the home unit cell in such a way that the cancellation in Eq. (4.7) is complete.

To see this, first consider the case of a disconnected Wannier band structure. As discussed earlier, we arrange the Wannier bands into a set of internally connected groups that are isolated from one another by gaps, and we work with Eqs. (4.12) and (4.14), which were formulated for this case. Clearly the arrangement of HW groups has to respect the \(z \leftrightarrow -z\) symmetry, so we can proceed as follows.

We organize the internally connected band groups into three collections as follows. First, we see whether there is a connected group centered at \(z = 0\), and if so, we call it the “origin-centered” group. In general, this can be done by listing all connected groups that pass through \(z = 0\). If the number of such groups is odd, the central one is the origin-centered group, and if not, there is no origin-centered group. A similar construction at \(z = c/2\) allows us to identify the “boundary-centered” group, if there is one. We then see whether there are any remaining HW groups lying between the origin-centered and boundary-centered groups (or between the corresponding gaps if these groups are absent). If so, we collect these, together with their \(z \rightarrow -z\) symmetry partners lying below the origin-centered group, into an “uncentered” collection of band groups. This leads to a unique procedure for defining the home unit cell as composed of the union of the origin-centered group, the boundary-centered group, and the uncentered collection. Figure 4.4(a) shows an example in which there are origin-centered and boundary-centered groups composed of two Wannier bands each, and an uncentered collection accounting for two more Wannier bands, for a system with six occupied valence bands.

We can also obtain the contributions of these three collections to various physical
properties. Let $C_{OC}$, $C_{BC}$, and $C_{UC}$ denote the total in-plane Chern numbers of the origin-centered, boundary-centered, and uncentered collections respectively, as computed from Eq. (4.16). We can also obtain the contribution of each collection to $\theta_{z\Omega}'$ in Eq. (4.12). We argued earlier that $\theta_{\Delta xy}' = 0$ in the presence of any axion-odd symmetry, so we denote the $\theta_{z\Omega}'$ contributions as simply $\theta_{OC}$, $\theta_{BC}$, and $\theta_{UC}$ respectively, where we have also dropped the prime for brevity. Then the total axion coupling is just

$$\theta_{cs} = \theta_{OC} + \theta_{BC} + \theta_{UC}.$$  \hspace{1cm} (4.28)

Now the contribution $\theta_{OC}$ from the origin-centered group takes the form given in Eq. (4.14), and this clearly vanishes in view of the cancellations between contributions at $z$ and $-z$ inside this Wannier band group (recall that $\bar{\Omega}_{nn}$ has the same sign in the canceling pair). Similar arguments imply that $\theta_{UC}$ also vanishes, since the groups below and above $z=0$ cancel in pairs. However, there is no such cancellation for the Wannier bands in the boundary-centered group, because their symmetry-mapped partners are

Figure 4.4: Sketches of hybrid Wannier band structures along a path in k-space that captures all the degeneracies of a system with an axion-odd symmetry that reverses $\hat{z}$.

(a) Disconnected Wannier band structure, decomposed into an origin-centered group (blue shading), a boundary-centered group (red shading), and an uncentered collection (yellow shading). The axion index depends on whether the Wannier bands in the red region have an odd Chern index, which in turn is determined by a counting of Dirac nodes; here $\theta = 0$ because the number of nodes in that region is even. (b) Connected Wannier band structure, decomposed into a thin slice centered at $z = c/2$ (red shading) and the remainder (blue shading). Here $\theta = \pi$ since there is an odd number of Dirac cones in the red region.
centered about $z = -c/2$ and are outside the chosen home unit cell (see Fig. 4.4(a), where the home cell consists of the union of the four shaded regions). Instead, each Wannier band at $(z, \mathbf{k})$ in the boundary-centered group has a partner in the same group at $(c - z, \pm g_\perp \mathbf{k})$. Again $\tilde{\Omega}_{nn}$ is identical at both locations, so when inserting $z_n$ into Eq. (4.14), we make no mistake if we treat both as located at the average location $z = c/2$. But then Eq. (4.14) becomes

$$\theta_{BC} = -\frac{1}{2} \int d^2k \sum_{n \in BC} \tilde{\Omega}_{nn} = -\pi C_{BC}$$

(4.29)

where Eq. (4.16) has been used. Since this is the only contribution, we have that $\theta_{cs} = \pi$ if and only if the Chern index of the boundary-centered group is an odd integer.

Now recall that we assumed that all bulk Chern numbers vanish in our material, so that $C_{BC} + C_{OC} + C_{UC} = 0$. Clearly $C_{UC}$ is even, since the uncentered collection is always composed of groups that contribute in pairs. Thus, $C_{BC}$ is odd if and only if $C_{OC}$ is odd. That is, the existence of an odd-Chern boundary-centered group also implies the existence of an odd-Chern origin-centered group. We are thus guaranteed to get the same result using the boundary-centered or origin-centered group in Eq. (4.29).

The same conclusion follows from the freedom to shift the origin by $c/2$ along $z$.

In some cases it may be possible to deduce the Chern number of a boundary- or origin-centered group by inspecting the nodal touching of Wannier bands. For example, suppose there are only two Wannier bands in the boundary-centered group, and that the locus of degeneracies between these bands consists only of some number $K$ of Dirac point nodes.\(^4\) In view of Eq. (4.26), these are typically found at $z = c/2$ in the Wannier direction, and at locations obeying $\pm g_\perp \mathbf{k} = \mathbf{k}$ in the 2DBZ. For example, they may occur at some of the four projected TRIM (PTRIM) for $g = E'$, or at other high-symmetry points or even at generic positions in the 2DBZ for other symmetries.

Because we are considering symmetries that preserve the Berry curvature while interchanging $z$ and $c - z$, and thus interchanging the two bands, we know that each Wannier band carries the same Berry flux $\Phi_0$. Now, if the symmetry is weakly broken in

\(^4\)To be a Dirac node, the dispersion $z_n(\mathbf{k})$ must be linear for all in-plane $\hat{\mathbf{k}}$ directions. This excludes quadratic and other higher-order nodal points from our discussion here.
such a way as to gap the Dirac nodes, then an additional Berry flux of $\pm \pi$ is transferred to each Wannier band at each of the nodes. However, this has to result in an integer Chern number, so we conclude that $\Phi_0$ must be 0 or $\pi \pmod{2\pi}$ if the number $K$ of Dirac nodes is even or odd, respectively. Recalling Eq. (4.29), this means that the axion $Z_2$ index is odd only if $K$ is odd. The same analysis can be applied to the origin-centered group.

This argument easily generalizes to the case that the number $M_{BC}$ of Wannier bands in the group is even, still assuming that the only touchings are Dirac-like. That is, the Chern number is odd if and only if the total number $K_{tot}$ of Dirac nodes is odd. In this counting procedure, a node involving an $L$-fold degeneracy is counted as $L/2$ Dirac nodes.

If the number $M_{BC}$ of Wannier bands in the group is odd, then we have to treat the central band separately. We let $j = (M_{BC} + 1)/2$ be the index of this band, and assume that it is regular enough to have a well defined Chern number $C_j$. Each nodal touching involving this band will also involve $L/2$ pairs of neighbors from among the remaining $(M_{BC} - 1)$ bands. These can be treated as before, providing $L/2$ Dirac nodes at one touching location, and a total number $K_{tot}$ when summed over the 2DBZ. The conclusion is that the total Chern index of the group, and thus the axion $Z_2$ index, is odd if and only if $C_j + K_{tot}$ is odd. This case is not quite as convenient, because it requires the evaluation of the Chern index of the central band, in addition to a simple counting of Dirac nodes.

In this analysis, we have assumed that the locus of contact between Wannier bands consists only of point Dirac nodes. If cases arise involving high-order (e.g., quadratic) point nodes, or nodal lines or regions, then the analysis given above would need to be reconsidered.

Essentially identical results involving the counting of Dirac nodes, and possibly the calculation of the Chern index of a central band, can be derived for the case of a connected Wannier band structure. To do so, we first establish a “nominal cell boundary” $z_{nom}(k)$ located near $z = -c/2$ as follows. Let $N$ be the number of Wannier bands passing through the point $(-c/2, \bar{\Gamma})$. If $N$ is odd, let $z_{nom}(k)$ be identified with
the central one of these bands; if \( N \) is nonzero and even, let it be the average of the central two bands; and if \( N = 0 \), let it be the average of the next lowest and next highest bands about \( z = -c/2 \). Then \( z_{\text{nom}}(k) \) is a surface centered at \(-c/2\) that respects the symmetries of the system and passes through the point \((-c/2, \bar{\Gamma})\).

If \( N \) is even, \( z_{\text{nom}}(k) \) lies between two Wannier bands. We assumed a connected band structure, so even if \( N = 0 \) these bands must touch at one or more nodes, which lie on the surface \( z_{\text{nom}}(k) \) by construction. If \( N \) is odd, then \( z_{\text{nom}}(k) \) is identified with a Wannier band, which again touches with the next higher band by assumption. In either case, let the cutting surface be

\[ z_{\text{cut}}(k) = z_{\text{nom}}(k) + \delta, \tag{4.30} \]

where \( \delta > 0 \) is a small vertical shift. With this convention, the unit cell consists of all portions of the Wannier bands lying between \( z_{\text{cut}}(k) \) and \( z_{\text{cut}}(k) + c \).

We first discuss the case of even \( N \). An example is sketched if Fig. 4.4(b); there the symmetry is such that \( z_{\text{nom}}(k) \) is perfectly flat at \( z = -c/2 \), and the unit cell consists of the union of the two shaded regions shown there. Regarding the degeneracies themselves, we let \( \epsilon \) to go to zero faster than \( \delta \to 0 \), so that the cutting surface avoids the DRs. Then \( \theta^{\prime\prime}_{z\Omega} \) reduces to the first term of Eq. (4.24). This in turn can be broken into two contributions: one from the blue region \( z_n(k) \in [z_{\text{nom}}(k) + \delta, z_{\text{nom}}(k) + c - \delta] \), and another from the red region \( z_n(k) \in [z_{\text{nom}}(k) + c - \delta, z_{\text{nom}}(k) + c + \delta] \), in Fig. 4.4(b). The first region is centered about \( z = 0 \), so all contributions cancel in pairs, while the latter region vanishes in the limit \( \delta \to 0 \), yielding a vanishing contribution to \( \theta^{\prime\prime}_{z\Omega} \). Thus, \( \theta^{\prime\prime}_{z\Omega} = 0 \). We argued earlier that the \( \theta_{\Delta xy} \)-type terms always vanish in the presence of axion-odd symmetries, so in fact the first two terms in Eq. (4.23) both vanish, and we are left with

\[ \theta_{cs} = \lim_{\delta \to 0} \phi_{\text{cut}} \quad (\text{even } N) \tag{4.31} \]

\[ ^5 \text{In the immediate vicinity of a Dirac node between a pair of Wannier bands, the Berry curvature arising from the coupling between the two crossing bands vanishes (although it diverges exactly at the node). Hence, the integrand in the first term of Eq. (4.24) is finite in the red region of Fig. 4.4(b) (excluding the DRs around the Dirac nodes), justifying the claim that the integral vanishes in the limit } \delta \to 0. \]
That is, the axion coupling is given by the sum of Berry phases of the vanishingly small loops of intersection of the Wannier bands with the cutting surface of Eq. (4.30) around the nodal points as $\delta \to 0$. If these are simple Dirac nodes lying on $z_{\text{nom}}(k)$, then we know that each such Berry phase is $\pi$, and it follows that the system is axion-odd if and only if the number of such Dirac nodes is odd, as in Fig. 4.4(b).

If the number $N$ of degenerate bands at $(-c/2, \Gamma)$ is odd, then arguments like those above yield

$$
\theta_{cs} = \lim_{\delta \to 0} \phi_{\text{cut}} - \pi C_j \quad (\text{odd } N),
$$

(4.32)

where $C_j$ is the Chern number of the central band. Again, if the nodes are simple Dirac nodes with their nodal points lying on $z_j(k)$, the system is axion-odd if and only if the sum of this central Chern index and the number of Dirac cones is odd. Note that we could equally well have chosen to work with a nominal cell boundary centered around $z=0$ instead of $z=-c/2$; as in the disconnected case, this follows from the freedom to shift the origin by $c/2$ along $z$.

**Symmorphic operation preserving $\hat{z}$** — Here we consider the case that $g$ preserves the sense of $\hat{z}$, so $g_\perp = g$ has to be either TR itself, a time-reversed rotation about the $z$ axis, or a simple reflection about a plane containing this axis, as summarized in Table 4.1. For the moment we assume there is no associated fractional translation along $\hat{z}$; that case will be considered in the next subsection. Now the Berry curvature contribution from an area element at $z$ on one Wannier band gets mapped onto one of opposite sign on the same sheet at the same $z$, giving canceling contributions to $\theta_{z\Omega}$.

In the disconnected case, this implies that $\theta = 0$ and the system is always in the axion-even phase, since any gap can be chosen as the primary one defining the unit cell. Conversely, if the system is axion-odd, then the Wannier bands must be connected.

For the connected case, any cutting surface $z_{\text{cut}}(k)$ that respects the in-plane symmetries will automatically give $\theta_{z\Omega}'' = 0$, since the same kind of cancellations occur. In this case, the axion $\mathbb{Z}_2$ index is just given by $\phi_{\text{cut}}$ in Eq. (4.23). We can arbitrarily choose one Wannier band $j$, make the cut at $z_j(k) + \delta$ for vanishingly small $\delta$, and count the total Berry phase of the small loops of intersection of the Wannier bands with $z_{\text{cut}}(k)$. In the case that these are simple Dirac nodes, the system is axion-odd if
and only if their number is odd.

_Nonsymmorphic operation preserving \( \hat{z} \)—_ Finally we again consider the \( \hat{z} \)-preserving case, but now assuming that the corresponding space-group operation involves a fractional translation \( c/m \) along \( \hat{z} \) for some integer \( m \). These can be glide mirrors, time-reversed half translations, and time-reversed screws, denoted as \{\( M_d|c/2 \), \( E'|c/2 \), \( C'_2|c/2 \), \( C'_3|c/2 \), \( C'_4|c/2 \), and \( C'_6|c/2 \), respectively._\(^6\)

For the moment, we focus on the operations that involve a half-lattice-vector translation, i.e., \{\( M_d|c/2 \), \( E'|c/2 \), \( C'_2|c/2 \), \( C'_3|c/2 \), \( C'_4|c/2 \), and \( C'_6|c/2 \). Then, according to Eq. (4.26), a Wannier band at \((z, \mathbf{k})\) always has a partner at \((z + c/2, \pm g_\perp \mathbf{k})\), with the minus sign applying when \( g_\perp \) involves TR. Moreover, all of the operations \( M_d \), \( E' \), \( C'_2 \), \( C'_3 \), \( C'_4 \), and \( C'_6 \) reverse the sign of the Berry curvature \( \tilde{\Omega} \), which is thus equal in magnitude but opposite in sign between these two partners.

If the band structure is disconnected, then there must be an even number of gaps. We choose a primary gap and divide the Wannier bands into two disconnected groups related to each other by Eq. (4.26): a “lower group” consisting of the first \( J/2 \) bands above the primary gap, and an “upper group” consisting of the rest. Consider a band \( n \) in the lower group; its contribution \( z_n(\mathbf{k})\tilde{\Omega}_{nn}(\mathbf{k}) \) to the integral in Eq. (4.14) can be paired with a contribution

\[
z_{n'}(\mathbf{k}')\tilde{\Omega}_{n'n'}(\mathbf{k}') = \left[z_n(\mathbf{k}) + \frac{c}{2}\right]\left[-\tilde{\Omega}_{nn}(\mathbf{k})\right]
\]

at \( \mathbf{k}' = \pm g_\perp \mathbf{k} \) in the upper group. The sum of these two contributions, when combined with the \(-1/c\) prefactor of Eq. (4.14), yields \( \tilde{\Omega}_{nn}(\mathbf{k})/2 \). Since we know that the \( \theta_{\Delta xy} \) term does not contribute, Eqs. (4.11-4.12) then yield

\[
\theta_{cs} = \frac{1}{2} \Phi_{LG} = \pi C_{LG},
\]

where \( \Phi_{LG} \) is the total flux of Berry curvature in the lower group, with \( C_{LG} \) being the corresponding Chern number (an integer). The system is thus axion-odd if and only if this Chern index is odd.

\(^6\)While \{\( C_3|c/2 \) and \( C_5|\pm c/6 \) are disallowed as elements of a nonmagnetic space group, since their cubes correspond to pure half-translations, the corresponding TR-composed operations are allowed in magnetic space groups.
To prepare for the connected case, let us return to the disconnected case for a moment. We start with a cutting surface \( z_{\text{cut}}^{(0)}(k) \) that lies entirely in the primary gap, so that the lower group has the total Berry flux \( \Phi_{\text{LG}} \) of Eq. (4.34), which we now relabel as \( \Phi_{\text{LG}}^{(0)} \). We then raise the cutting surface to some chosen new location \( z_{\text{cut}}(k) \) that cuts through some of the bands at the bottom of the lower group. (We also insist that \( z_{\text{cut}}(k) \) avoids any DRs.) The new lower group now extends from \( z_{\text{cut}}(k) \) to \( z_{\text{cut}}(k) + c/2 \). As a result, the total Berry flux \( \Phi_{\text{LG}} \) in this group changes for two reasons: because of the omission of terms below \( z_{\text{cut}}(k) \), which is compensated by the addition of a term \( \phi_{\text{cut}} \); and by the addition of new contributions near \( z_{\text{cut}}(k) + c/2 \), which is compensated by the addition of the same \( \phi_{\text{cut}} \) because of the sign reversal of \( \tilde{\Omega}_{nn} \). Thus, we find that

\[
\Phi_{\text{LG}}^{(0)} = \Phi_{\text{LG}} + 2\phi_{\text{cut}}.
\] (4.35)

The quantity on the right side of this equation is \( 2\pi \) times an integer, even though the individual terms are not, and the system is axion-odd if and only if this integer is odd.

Finally we argue that this same formula applies to the connected case, since following in the spirit of the discussion in Sec. 4.1.2, we can imagine that we introduce a perturbation that opens a gap, place the cutting surface there, raise the cutting surface, and then close the gap. A concise representation of the final result is

\[
\theta_{cs} = \frac{1}{2} \Phi_{\text{LG}} + \phi_{\text{cut}}.
\] (4.36)

This formula should be correct for any choice of cutting surface, as long as it avoids the DRs.

In the case of Dirac cones connecting a pair of bands, we can choose a cutting surface just above the average of these bands, so that \( \phi_{\text{cut}} \) could again be obtained as \( \pi \) times the number of such Dirac cones. However, unlike the preserving and reversing the \( \hat{z} \)-direction cases, here we cannot avoid doing an explicit calculation of the total Berry flux in half the unit cell. In this sense, Eq. (4.36) is not quite as convenient as our earlier connected-case formulas.

We now consider the operations involving smaller fractional translations, namely \( \{C_3'|c/3\} \), \( \{C_6'|c/3\} \), \( \{C_4'|c/4\} \), \( \{C_3'|c/6\} \), and \( \{C_6'|c/6\} \). (We omit left-handed screws, since they behave in the same way as right-handed ones.)
We start by assuming a disconnected band structure, and consider the example of $\{C'_4|c/4\}$. The number of gaps must be a multiple of four; we choose one such gap to define the unit cell in the usual way. Now $\{C_2|c/2\}$, which is the square of $\{C'_4|c/4\}$, is also in the symmetry group. It simply translates the Wannier bands by a half lattice vector along $\hat{z}$ with an extra $C_2$ rotation without affecting the value of the Berry curvature on a given patch of the Wannier band, which is just carried along to its new location. From the point of view of Eqs. (4.7) or (4.14) for the contribution to $\theta_{cs}$, the rotation component is irrelevant, since it does not affect the $z_n$ or $\Omega_{nn}$ value. Thus, for our purposes we can think of $\{C_2|c/2\}$ as defining a smaller “unit subcell” of height $c' = c/2$, and we can compute $\theta_{cs}$ by focusing on just one subcell containing $J' = J/2$ Wannier bands.

For the case of isolated bands, for example, Eq. (4.7) becomes

$$\theta \Omega = -\frac{1}{c'} \int d^2 k \sum_{n=1}^{J'} z_n \Omega_{nn} \quad (4.37)$$

(the fact that we count half as many bands is compensated by the factor of two from the replacement of $c$ by $c'$ in the prefactor), and a similar modification would apply to Eqs. (4.10) and (4.12) in the case of composite groups. Now the action of $\{C'_4|c/4\}$ in the subcell is entirely analogous to the action of $\{C_2|c/2\}$ in the full cell; this is one of the cases that we studied above, and the same conclusions apply. That is, we conclude that $\theta_{cs}$ is given by Eq. (4.34), where $C_{LG}$ is now interpreted as the total Chern number coming from the bottom half of the subcell, i.e., from the first $J/4$ bands. The case of a connected band structure under $\{C'_4|c/4\}$ is handled using the same analogy to the $\{C'_4|c/2\}$ case, and a formula like Eq. (4.36) applies.

The same strategy can be applied to reduce the remaining operations to previously studied ones in a similar way. All of these remaining operations, namely $\{C'_3|c/3\}$, $\{C'_6|c/3\}$, $\{C'_3|c/6\}$, and $\{C'_6|c/6\}$, have the property that their fourth power is (modulo full translations) just a simple 3-fold screw, which divides the unit cell into three subcells. Thus, we can restrict our attention to a subcell of height $c' = c/3$ containing $J' = J/3$ bands.
For the first two operations, \( \{C'_3|c/3\} \) and \( \{C'_6|c/3\} \), there are no remaining fractional translations within the subcell. Their third powers correspond to \( E' \) and \( C'_2 \) respectively, so in these cases the analysis of the subcell is analogous to the symmorphic case discussed in the preserving \( \hat{z} \) direction, and the conclusions found there apply. Specifically, \( \theta_{cs} \) is trivial in the disconnected case, and a counting of Dirac nodes can provide the value of \( \theta_{cs} \) in the connected case.

For the last two operations, \( \{C'_3|c/6\} \) and \( \{C'_6|c/6\} \), the subcell of height \( c/3 \) is subdivided by a further half translation of \( c/6 \). The situation is similar to the \( \{C'_4|c/4\} \) case discussed above, except that now the subcell is of size \( c/3 \) instead of \( c/2 \). This subcell has lower and upper portions containing \( J/6 \) Wannier bands each, related to each other by an operation containing TR. Thus, the analysis of the subcell in these cases is analogous to the that of the full cell in the nonsymmorphic \( \{E'|c/2\} \) and \( \{C'_2|c/2\} \) cases respectively. In the disconnected case, Eq. (4.34) applies with \( C_{LG} \) interpreted as the Chern number of the first \( J/6 \) bands, and in the connected case a formula like Eq. (4.36) once again applies.\(^7\)

**Summary** — In summary, for the case of a disconnected band structure, the rules for determining the axion index are quite simple. For a \( z \)-reversing symmetry operator, we test for the presence of a boundary-centered group with an odd Chern number, or equivalently, an origin-centered one; if present, the system is axion-odd. For symmetry operators that preserve the sign of \( \hat{z} \), we can consider three cases. If there is no associated fractional translation along \( \hat{z} \), the system is forced to be axion-trivial. For the case of a half translation, the system is axion-odd if and only if the total Chern number in a half unit cell is odd. For smaller fractional translations, we can define a subcell that tiles the full cell under simple screw rotations, and apply the earlier analysis to the subcell.

The rules for the connected case are more complicated, but in the simplest case that the connection is via Dirac nodes, the results can be expressed in terms of a counting of Dirac nodes, and may also require the calculation of the total Berry flux of a band

\(^7\)Alternatively, Eq. (4.34) or Eq. (4.36) can be applied directly in the full cell using the previous results for \( \{E'|c/2\} \) or \( \{C'_2|c/2\} \), the third powers of \( \{C'_3|c/6\} \) and \( \{C'_6|c/6\} \) respectively.
or subgroup of Wannier bands in the unit cell.

Note that we may have some choice regarding the crystallographic direction along which to perform the wannierization. In the case of $I$ or $E'$, any convenient primitive reciprocal direction will do. In the case of a mirror, we may choose either an axis normal to or lying in the mirror plane, applying the analysis of $\hat{z}$-reversing or $\hat{z}$-preserving respectively. (If $\theta_{cs} = \pi$, the Wannier bands are necessarily connected for the choice of wannierization in the mirror plane.) The case of $C'_2$ rotations provides a similar choice of options, i.e., the HW construction can be done either normal to the $C_2$ axis ($\hat{z}$-reversing) or along it ($\hat{z}$-preserving). Of course, there may be more than one axion-odd symmetry in the magnetic point group, in which case one also has a freedom to choose which operation to select. Thus, it may often be possible to simplify the determination of the axion index by an appropriate choice of symmetry and setting.

### 4.3 Bulk-boundary correspondence

Even though the axion coupling $\theta_{cs}$ is a bulk quantity, it has important implications for surfaces. A given surface facet of an axion-odd insulator will either be metallic with an odd number of Dirac cones and a vanishing surface AHC, or it will be gapped with a half-integer surface AHC of $(N + 1/2)e^2/h$. Furthermore, if all surfaces are gapped, then $\theta_{cs} = \pi$ implies a higher-order bulk-boundary correspondence. That is, both bulk and surfaces are gapped, but one-dimensional chiral channels are present.

Previous work has shown that there is often a close relationship between the Wannier band structure in a given crystallographic direction, and the surface band structure for a surface facet normal to that direction. For example, the flow of the Wannier bands for a weak or strong TI in a 3D TR-invariant insulator is reflected in the presence and character of the surface energy bands on the corresponding surface facet. Thus, it is of interest to explore the relationship between the Wannier band structure and the corresponding surface band structure in axion-odd insulators.

In Secs. 4.3.1 and 4.3.2 we address two closely related questions concerning spectral flow in an axion-odd insulator. First, we identify the symmetry conditions under which
topologically protected metallic states must exist on a given surface. Conversely, when are insulating surfaces allowed to appear? We then discuss the conditions under which the flow of the bulk Wannier bands in a given direction is topologically protected. Conversely, when can an axion-odd insulator have a gapped Wannier band structure? In Sec. 4.3.3 we discuss the identification of axion-odd insulators as second-order topological insulators, discuss the conditions on the occurrence and the number of chiral channels on a given hinge, comment on their robustness, and make connections with previous literature on second-order topological phases.

4.3.1 Protected flow of the surface energy bands

To answer the first question posed above, we ask whether the surface magnetic point group contains any axion-odd symmetry operations. If so, the surface AHC would be forced to vanish (see below). However, this is inconsistent with the half-integer surface AHC that, according to Eq. (2.78), must occur for any insulating surface of our assumed axion-odd bulk insulator. Thus, under these conditions the surface is necessarily metallic. In this case, Eq. (2.78) becomes

\[ \sigma_{\text{AHC}}^{\text{surf}} = \frac{e^2}{\hbar} \frac{\phi - \theta_{cs}}{2\pi} \mod e^2/h, \]  

(4.38)

where the extra phase angle \( \phi \) is the Berry phase summed over all the Fermi loops in the surface BZ. Since each Dirac cone contributes a Berry phase of \( \pi \), a topological metallic surface must have an odd number of them to cancel the \( \theta_{cs} = \pi \) contribution from the bulk.

If no axion-odd symmetries are present at the surface, then the surface is allowed to be insulating. Of course, whether it is really insulating or not will depend on details of the surface electronic structure for the particular surface termination, but the bulk topology no longer requires a metallic state.

One immediate consequence is that if TR itself is a symmetry of the bulk, and also of the surface of interest, then this surface must be metallic. This is the well-known statement that strong TIs have metallic surface states on all surfaces, unless TR is somehow broken on the surface. By contrast, if inversion is the only axion-odd
symmetry protecting the axionic topology of the bulk, then all surfaces are allowed to be insulating, because no surface preserves inversion symmetry.

More complicated situations can be analyzed as follows. For a given facet on the surface of an axion-odd crystal, we first reduce from the 3D bulk magnetic space group to the 2D magnetic space group that describes the highest symmetry that the surface could possibly have. This will consist of all operations of the 3D group that leave the surface normal (chosen as $\hat{z}$) invariant, and that involve no translation (either by a full or a fractional lattice vector) along $\hat{z}$; the axion-odd operations among these, if any, are of the $z$-preserving type. We then construct the surface magnetic point group in the usual way, by listing the point operations appearing in the space group. Finally, we omit any operations that are not axion-odd. The remaining operations that may be present are the $z$-preserving operations in Table 4.1: TR itself, the $C'_{n}$ time-reversed rotations about $\hat{z}$ ($n = \{2, 3, 4, 6\}$), and reflections $M_{d}$ about a plane whose unit normal lies in the surface. All these symmetries flip the sign of the surface AHC (a pseudovector pointing along $\hat{z}$), forcing it to vanish. Hence, if any of these symmetries are present at the surface, then the surface must be metallic, with $\phi = \pi$ in Eq. (4.38).

Of course, any given surface may have lower symmetry than the one allowed by the above considerations. For example, a bulk-allowed $M_{d}$ mirror symmetry may be spontaneously broken by the formation of a symmetry-lowering surface reconstruction, or TR may be broken at the surface by the spontaneous appearance of magnetic order. In general, if any of the axion-odd symmetries survive at the surface, it must be metallic; otherwise it can be insulating.

We mentioned earlier that if inversion is the only axion-odd bulk symmetry, then all surfaces are allowed to be insulating. One may ask whether this is also true for any other symmetries. The answer is yes. Consider, for example, the case that an improper rotation $S_{3}$, $S_{4}$, or $S_{6}$ about the $z$ axis is a symmetry; then for any facet normal $\hat{n}$, the component of $\hat{n}$ on the $(x, y)$ plane is rotated by the symmetry, and the $z$ component is reversed, so there are no surfaces that obey this symmetry. The time-reversed screw rotations also have this property. This time they do preserve $\hat{n}$ for surfaces normal to the rotation axis, but the fractional translation that comes with the screw operation is
always inconsistent with such a surface.

4.3.2 Protected flow of the bulk Wannier bands

As we have seen, a surface of unit normal \( \hat{z} \) on an axion-odd insulator is required to be metallic if any bulk axion-odd symmetries of the \( z \)-preserving type are preserved at the surface. Assume that such a high-symmetry metallic surface has been prepared. Following Fidkowski et al. [39], Neupert and Schindler [85], the surface spectrum can be continuously deformed into the bulk Wannier spectrum obtained by wannierizing along \( \hat{z} \). Briefly, the actual surface band structure is connected to a model surface obtained by energetic flattening and spatial truncation, and then the abrupt truncation is replaced by a crossover region whose width is allowed to diverge. In this limit, these authors show that the Wannier band structure is recovered. Since the surface spectrum flows by hypothesis, the Wannier spectrum must flow as well.

This argument, based on the bulk-boundary correspondence, reproduces the conclusions of the \( \hat{z} \)-preserving case, where purely bulk considerations led to the conclusion that the Wannier bands must be connected if a \( z \)-preserving axion-odd symmetry is present. These considerations also imply that the flow should occur via an odd number of Dirac nodes between adjacent Wannier bands.

We emphasize, however, that the bulk-boundary argument does not necessarily work in reverse. That is, if the surface band structure does not flow, this does not necessarily imply the same for the bulk Wannier bands. Sometimes a \( z \)-reversing or \( z \)-nonsymmorphic symmetry can lead to flow of the Wannier bands in a “fragile” sense. By this we mean that the flow can be destroyed, without breaking any axion-odd symmetries, by adding some weakly-coupled trivial bands to the valence manifold; this behavior was demonstrated in Wieder and Bernevig [139] for the case of inversion symmetry. This is again consistent with the analysis of \( \hat{z} \)-reversing case, where both connected and disconnected Wannier band structures were found to be compatible with \( \theta_{cs} = \pi \). This will be illustrated for the case of a pyrochlore model with inversion symmetry, and for the case of a model with glide mirror symmetry. Since \( z \)-reversing and \( z \)-nonsymmorphic operations are never symmetries of the surface, there is never
any such fragile protection of the flow of the surface energy bands in these cases.

### 4.3.3 Higher-order topology

Here we revisit the bulk-boundary correspondence for an axion-odd insulator, now relaxing the requirement that the surface facet in question is itself invariant under the symmetry, as was assumed in Sec. 4.3.1. We consider a macroscopic crystallite, and only require that the boundary as a whole is left invariant under an axion-odd symmetry.

Under these conditions, facets that are mapped to themselves under the symmetry are necessarily metallic, but those that are not can be insulating. When that happens, an insulating facet gets mapped into a sequence of $n-1$ other facets for a symmetry operator of order $n$; we take these to comprise an “orbit.” All facets in the orbit have the same magnitude of surface AHC, but the sign alternates from facet to facet. There could be more than one orbit in general. Whenever insulating facets with different values of surface AHC meet, the difference must be an integer multiple of $e^2/h$, with that integer corresponding to the number of chiral modes propagating along the connecting hinge [109]. Since the sign of the AHC cannot possibly be uniform on all facets, some hinge modes are necessarily present.

Such axion-odd crystals thus constitute examples of second-order TIs [5, 6, 36, 61, 68, 106, 107, 110, 122, 134], i.e., crystals whose bulk topology in $d=3$ dimensions gives rise to protected boundary states of dimension $d-2=1$. For example, consider an axion-odd insulator with $C'_4$ symmetry [106], and take a crystallite with a tetragonal shape. The top and bottom surfaces are required to be metallic because they are mapped onto themselves by $C'_4$ symmetry. Instead, the four side surfaces can be insulating, and we assume they are. In that case they form an orbit in which the AHC of adjacent side facets alternates between $\pm(N + 1/2)e^2/h$ for some fixed integer $N$. This results in $2N + 1$ chiral modes on each of the four vertical hinges, whose chiralities change sign from one hinge to the next.

A similar story applies to the case of a tetragonally-shaped crystallite invariant under not $C'_4$ but $S_4$ symmetry [122]. In this case no facet is left invariant under the axion-odd symmetry $S_4$, so that all six facets are allowed to be insulating. The
surface AHC alternates between \(\pm (N_1 + 1/2) e^2/h\) on the side-surface orbit, and between \(\pm (N_2 + 1/2) e^2/h\) on the top-and-bottom orbit, where the integers \(N_1\) and \(N_2\) depend on the surface preparation. The pattern of hinge modes will depend on the integers \(N_1\) and \(N_2\), but no choice of these integers can prevent the appearance of hinge modes.

An *intrinsic* [42, 119, 120] second-order TI is one whose bulk topology is such that metallic hinge modes are required if the entire crystallite, including the crystal termination, is invariant under the topology-protecting symmetries. It follows that the axion-odd insulators discussed here belong to this class. By contrast, the classification of *extrinsic* [42, 119, 120] second-order TIs defines topological equivalence based on continuous deformations that preserve bulk and surface gaps [42]. In this context, we note that the chiral hinge modes on axion-odd insulators are robust against small deformations of the Hamiltonian that preserve bulk and surface gaps, even if they break the axion-odd protecting symmetries [109]. We also note that in some circumstances, chiral edge channels can appear not only on hinges, but also on steps and domain-walls within a single facet [79, 126]. On the other hand, sufficiently strong surface deformations can always remove the chiral channels, leading to the topological magnetoelectric effect. This can be arranged, for example, by appropriately “decorating” the surfaces of the crystallite with quantum anomalous Hall layers in such a way that the boundary has a uniform half-integer AHC [126]. In any case, it is clear that the Wannier band structure cannot predict when and where extrinsic hinge states will appear, since it is a purely bulk construction.

### 4.4 Special cases and numerical tests

In this section we consider four axion-odd symmetry operations that deserve special emphasis: TR (i.e., \(E'\)); inversion \(I\); mirror (either \(M_z\) or \(M_d\), depending on the wannierization direction); and glide mirror \(\{M_d|c/2\}\). In each case we shall introduce an elementary tight-binding model that can be used to illustrate some of the behaviors expected from theory.
4.4.1 Time-reversal symmetry

General considerations — Here we consider the case that TR itself, $E'$, is a symmetry of the system. This is a $z$-preserving symmetry and the conclusions given in that paragraph apply directly.

We first consider the case of spinor electrons, so that $(E')^2 = -1$. As is well known, all energy bands of a 2D system appear as Kramers pairs at all four TRIM. In a similar way, the Wannier bands of a 3D system all appear as Kramers degeneracies at the four PTRIM, as discussed in Taherinejad et al. [112]. Therefore, any given Wannier band $j$ is either Kramers degenerate with band $j - 1$, or with band $j + 1$, at each of the four PTRIM. Let us call these “down-touchings” and “up-touchings” respectively; their numbers are $N_d(j)$ and $N_u(j) = 4 - N_d(j)$ respectively. The next band clearly has $N_d(j + 1) = N_u(j)$, etc. Since $N_u(j + 1) = 4 - N_d(j)$, there are three cases to consider:

- $N_d$ alternates between 0 and 4 as bands are counted. The Wannier bands are thus “glued together” in pairs, and are generically gapped from the next higher or next lower pair. From this it follows that the axion $Z_2$ index is trivial. As discussed in Taherinejad et al. [112], it also follows that the $Z_2$ indices are trivial on the four TR-invariant planes $k_x = 0$, $k_x = \pi$, $k_y = 0$, and $k_y = \pi$ in the 3DBZ. This system is either topologically trivial, or it is a weak topological insulator with indices $(0;001)$, i.e., equivalent to a stack of quantum spin Hall (QSH) layers along $z$.

- $N_d = 2$ for all bands. The Wannier band structure is connected; there are no gaps. These systems correspond to weak topological insulator configurations, this time corresponding to a stacking of QSH layers along $x$ or $y$ or in the diagonal $xy$ direction. According to the discussion for the $z$-preserving case, we can choose a cutting surface just above any chosen band; the two Dirac cones cut by this surface contribute an even integer times $\pi$, and the system is again axion-trivial.

- $N_d$ alternates between 1 and 3 as bands are counted. The Wannier band structure

---

\textsuperscript{8}As an exception, if other symmetries are also present, these could induce additional degeneracies between Wannier bands, typically at high-symmetry points other than the TRIM.
is again connected. A cutting surface is again chosen just above one Wannier band; since there will be 1 or 3 Dirac cones cut by this surface, each carrying a Berry phase of $\pi$, the system is a strong topological insulator and it is axion-odd.

These conclusions are consistent with the well-known properties of TR-protected topological insulators, as discussed, for example, in Taherinejad et al. [112].

For the case of scalar particles, $(E')^2 = +1$, the Kramers degeneracies are not enforced, and the Wannier band structure will generically be disconnected, in which case $\theta = 0$. In case a connected Wannier band structure is enforced by degeneracies associated with additional symmetries, the Dirac node counting procedure described at the end of $\hat{z}$-preserving paragraph of Sec. 4.2.2 should still apply.

**Fu-Kane-Mele model** — To illustrate how the trivial, weak-TI, and strong-TI phases manifest themselves in the HW representation, we use the FKM model [41]. In this model, spinor $s$-like orbitals interact with spin-dependent and spin-independent hoppings on a diamond lattice of conventional lattice constant $a$. The Hamiltonian contains two terms,

$$H = \sum_{\langle i j \rangle} t_{ij} c_i^\dagger c_j + \lambda_{SO} \sum_{\langle\langle i j \rangle\rangle} i c_i^\dagger g_{ij} \cdot \sigma c_j.$$

The first term describes spin-independent first-neighbor hoppings, denoted as $\langle i j \rangle$; the hopping amplitudes are $t_{ij} = t_0$ for all bonds except those aligned along $[111]$, for which $t_{ij} = t_0 + \Delta t$. The second term, which models the effect of spin-orbit coupling, consists of spin-dependent second-neighbor hoppings denoted as $\langle\langle i j \rangle\rangle$; $\sigma = (\sigma_x, \sigma_y, \sigma_z)$ are the Pauli matrices, and $g_{ij} = d_{il} \times d_{lj}$ where $d_{il}$ and $d_{lj}$ are the bond vectors of the first-neighbor pairs making up the second-neighbor bond. We set $t_0 = 1$ and $\lambda_{SO} = 0.125$, and adjust $\Delta t$ according to the phase diagram

$$\Delta t \in \begin{cases} (-\infty, -4), & \text{Trivial insulator} \\ (-4, -2), & \text{Strong TI} \\ (-2, 0), & \text{Weak TI} \\ (0, 2), & \text{Strong TI} \\ (2, \infty), & \text{Trivial insulator} \end{cases}$$
to put the system in a trivial-insulator, weak-TI, or strong-TI phase at half filling [112].

Figure 4.5: Wannier band structures for the FKM model in a trivial-insulator phase with $\Delta t = 2.5$ (a,d), a weak-TI phase with $\Delta t = -1.0$ (b,e), and a strong-TI phase with $\Delta t = 1.0$ (c,f). In (a-c) the wannierization is along the (111) direction, denoted as $\hat{z}'$, with lattice periodicity $c' = a\sqrt{3}/3$; in (d-f) we wannierize along (001), with lattice periodicity $c = a/2$. The bands are plotted along high-symmetry lines connecting the four PTRIM in the 2DBZ.

For our numerical tests we pick a trivial insulator with $\Delta = 2.5$, a weak TI with $\Delta t = -1.0$, and a strong TI with $\Delta t = 1.0$. Their Wannier band structures are shown in Fig. 4.5; in the top panels the wannierization is along (111), and in the bottom panels it is along (001).

As explained earlier, in a trivial insulator the number $N_d$ of Kramers down-touchings at the PTRIM is expected to alternate between 0 and 4 irrespective of the wannierization direction, and this is indeed what is observed in the left panels of Fig. 4.5.

As for the middle panels of Fig. 4.5, we note that a weak TI can always be thought of as a stack of QSH layers in some direction determined by the weak indices [41]. In this example the weak-TI phase corresponds to a stack of QSH layers along (111), which should not come as a surprise since we have weakened the hoppings along [111]. Hence, when wannierizing along (111) we expect a gapped spectrum with $N_d$ alternating between 0 and 4, as seen in Fig. 4.5(b). On the other hand, for other wannierization directions we expect a connected spectrum with $N_d = 2$ for all bands, as seen in
Fig. 4.5(e).

Finally, the behavior of the right panels of Fig. 4.5 is as predicted for a strong TI: the Wannier band structure is necessarily connected, with $N_d$ alternating between one and three.

### 4.4.2 Inversion symmetry

*General considerations* — Next we consider the case that a simple inversion $I$ is a symmetry of the system, so that it is denoted as an “axion insulator” if it is axion-odd. This is an example of a $z$-reversing axion-odd symmetry, and the conclusions derived in Sec. 4.2.2 apply. To summarize, we found that in the disconnected case, $\theta_{cs} = \pi$ if and only if there is a boundary-centered group and its Chern number $C_{BC}$ is odd. (The same applies to the origin-centered group.) Unlike TR, the presence of inversion symmetry does not require an axion-odd insulator to have a connected Wannier band structure [126, 139], as discussed in Sec. 4.3.2. Regardless of whether it does or not, we can frequently deduce the $Z_2$ index from a node-counting argument, e.g., by focusing on the set of Wannier bands that touch $z = -c/2$ at $\bar{\Gamma}$. If the number of bands in this group is even, we place a cutting surface infinitesimally above the average of the central two bands, and conclude that the $Z_2$ index is odd if and only if the number of Dirac nodes sliced in this way is odd. In case the number of Wannier bands is odd, then we need information about the Chern number of the central band in addition. The same considerations apply to an analysis centered around $z = 0$. We note in passing that the work of Alexandradinata et al [2] provides a complementary view of the Wannier (Wilson-loop) band structure for 2D insulators.

The axion $Z_2$ index can also be obtained from a very different approach based on parity counting. In a path-breaking work, Turner et al [121]. derived a set of rules for deducing many topological properties of a centrosymmetric crystalline material based on a counting of the odd-parity eigenvalues of the occupied Bloch states at the eight TRIM in the 3DBZ. They showed that the system can be insulating only if the total number of odd-parity states is an even integer, and moreover it is axion-even or axion-odd depending whether this number is of the form $4n$ or $4n + 2$, where $n$ is an integer.
In the Appendix, we carry this analysis over to the Wannier band structure. Here, the original eight TRIM of the 3DBZ project onto four PTRIM in the 2DBZ. At each of these $k$, the Wannier bands can be classified as being of even or odd parity at $z=0$; even or odd parity at $z = c/2$; or appearing in pairs at $\pm z$. We develop counting rules inherited from the 3D parity counting of the Bloch states. Consistent with the analysis above, we find that the axion $\mathbb{Z}_2$ index can often be determined directly from an inspection of the Wannier band structure, even without evaluating the parities of the HW states. In some cases, however, some parity eigenvalues do need to be evaluated, allowing a determination of whether the boundary-centered group has an odd Chern number as discussed above. The details of this analysis are deferred to the Appendix.

**Pyrochlore model** — For the case of inversion symmetry we borrow an illustrative class of centrosymmetric tight-binding models discussed in Varnava and Vanderbilt [126]. These models consist of one pair of spinor basis orbitals per site of the pyrochlore lattice, as might be used to describe the $J_{\text{eff}} = 1/2$ manifold of the pyrochlore iridates. The Hamiltonian reads

$$ H = t \sum_{\langle ij \rangle} c_i^\dagger c_j + \lambda \sum_{\langle ij \rangle} i\sqrt{2}c_i^\dagger \hat{g}_{ij} \cdot \sigma c_j + \Delta \sum_i \hat{n}_i \cdot \sigma c_i^\dagger c_i, \quad (4.40) $$

where the spin indices are suppressed. Just as for the FKM model, the first two terms describe spin-independent and spin-dependent first-neighbor hoppings, with amplitudes $t$ and $\lambda$ respectively. In the latter, $\hat{g}_{ij}$ is a unit vector normal to both the hopping and the potential gradient directions; see Varnava and Vanderbilt [126] for details. The third term is a Zeeman term responsible for breaking TR symmetry, with the vectors $\hat{n}_i$ ($i = 1, 2, 3, 4$) describing the directions of the magnetic moments on the pyrochlore lattice sites. We choose the all-in-all-out magnetic configuration [126], and consider the model at half filling.

Whether the system is axion-even or axion-odd can easily be determined using the parity criteria for inversion-symmetric insulators [121]. That is, the system is even or odd if the total number of odd-parity Bloch states at the TRIM is of the form $4n$ or $4n + 2$ respectively, where $n$ is an integer. Table 4.2 shows how the odd-parity states are distributed among the eight TRIM for three representative choices of
Table 4.2: Number of odd-parity Bloch states at the eight TRIM for three representative cases: “Trivial” phase with \((t, \lambda, \Delta) = (1.0, 0.1, 3.0)\); “Axion I” phase with \((t, \lambda, \Delta) = (1.0, 0.1, 0.6)\); and “Axion II” phase with \((t, \lambda, \Delta) = (-1.0, -0.1, 0.6)\). The TRIM are labeled as in Wan et al. [130].

<table>
<thead>
<tr>
<th>Phase</th>
<th>(\Gamma)</th>
<th>(X, Y, Z)</th>
<th>(L (\times 3))</th>
<th>(L')</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trivial</td>
<td>0</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>Axion I</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>Axion II</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>14</td>
</tr>
</tbody>
</table>

parameters (see the caption) corresponding to different regions of the phase diagram. Their total number is 12, 10, and 14, indicating axion-even, axion-odd, and axion-odd topology, respectively. We refer to them below as the trivial, Axion I, and Axion II cases, respectively.

We present the corresponding Wannier bands in Fig. 4.6. Along (111) the pyrochlore lattice consists of alternating triangular and kagome layers containing one and three atoms, respectively. This is reflected in the Wannier band structure of the trivial phase in Fig. 4.6(a), since in that phase electrons are localized on the atomic sites. Similarly, along (001) we have a stack of tetragonal layers rotated by 45° with respect to each other, and in Fig. 4.6(d) we see that the OC and BC groups have two Dirac nodes each, implying \(\theta=0\).

In the axion-insulator phase, the Wannier band structure can be connected or disconnected depending on the choice of parameters, as illustrated in the middle and right panels of Fig. 4.6. The two axion phases differ by four odd-parity states (see Table 4.2), consistent with both being axion-odd. The fact that there is flow in the case of Axion I state, but not for Axion II, is another illustration of the kind of fragile protection that was discussed in Sec. 4.3.2. For example, it is evident that the addition of a pair of trivial flat Wannier bands near \(z=\pm0.25c\) would convert the hybrid Wannier band structure in Fig. 4.6(b) into a disconnected one after hybridization with the crossing bands would occur.

In the connected Axion I case, the number of Dirac nodes at the nominal cell
boundary is an odd integer, as per Eq. (4.31). Indeed, in Fig. 4.6(b) there are five nodes at \( z = c/2 \) and three at \( z = 0 \), while in Fig. 4.6(e) there are three nodes at both \( z = c/2 \) and \( z = 0 \).

Figure 4.6: Wannier band structures for the pyrochlore model, wannierized along (111) (denoted \( \hat{z}' \)) in (a-c), and along (001) in (d-f). (a,d) Trivial insulator; (b,e) Axion I phase; (c,f) Axion II phase (see caption of Table 4.2).

For the disconnected band structures in the left and right panels of Fig. 4.6, we showed in Eq. (4.29) that \( \theta_{cs} = 0 \) or \( \pi \) depending on whether the Chern index of the boundary-centered group is even or odd. While it is possible in principle to compute Berry curvatures and Chern numbers from the bulk Wannier bands [86, 111], here for convenience we carry out the analysis based on a slab configuration chosen thick enough that the central Wannier bands are bulk-like. In particular, we construct slabs with a thickness of ten unit cells along \( \hat{z}' \parallel (111) \). Slab HW states maximally-localized along \( \hat{z}' \) are obtained as eigenstates of the projected position operator \( P_k \hat{z}' P_k \) [64, 77, 86]. For a sufficiently thick slab, the ones located far from the surfaces are indistinguishable from the bulk HW states. We then organize the corresponding eigenvalues – the Wannier bands of the slab – into origin-centered and boundary-centered groups, i.e., centered at integer and half-integer \( z'/c' \), respectively. Finally, we calculate explicitly the Chern index of each group of Wannier bands using Eq. (4.29). In Fig. 4.7 we confirm that in the disconnected Axion II phase the Chern index of the boundary-centered groups is
an odd integer \((C = -1)\), while in the trivial phase it is even \((C = 0)\).

For the Axion II phase, we could have reached the same conclusion from either Fig. 4.6(c) or (f) by counting Dirac nodes, since the origin centered and boundary centered groups have two Wannier bands each. For the trivial phase this strategy cannot be used for the case of 3+1 bands in Fig. 4.6(a), but it does does apply to the (001) wannierization with 2+2 bands in Fig. 4.6(d), confirming the trivial topology.

![Chern index for connected groups of slab Wannier bands of the pyrochlore model](image)

Figure 4.7: Chern index for connected groups of slab Wannier bands of the pyrochlore model, plotted as a function of their average positions \(\bar{z}'\). Red and blue markers correspond to groups with integer and half-integer \(\bar{z}'/c'\) respectively. We consider slabs with a thickness of 10 unit cells along (111) in (a) the trivial phase of Fig. 4.6(a), and (b) the axion II phase of Fig. 4.6(c); in the axion phase, the top layer was “exfoliated” to remove unwanted nontopological surface states. In the trivial phase, both the groups with integer \(\bar{z}'/c'\) – one band – and those with half-integer \(\bar{z}'/c'\) – three bands – have zero Chern index. In the axion II phase both groups comprise two bands each, and far from the surfaces their Chern indices are \(-1\) and \(+1\) respectively.

### 4.4.3 Mirror symmetry

*General considerations* — In Sec. 4.2.1 we considered two types of mirrors relative to a given wannierization direction \(\hat{z}\). When \(\hat{z}\) lies on the mirror plane, the mirror is of type \(M_d\), and is a symmorphic \(z\)-preserving operation as discussed in Sec. 4.2.2. When \(\hat{z}\) is perpendicular to the mirror plane, we have an \(M_z\) mirror reflection, a \(z\)-reversing axion-odd symmetry of the kind presented in Sec. 4.2.2.
The case of $M_d$ is similar to that of TR, while the case of $M_z$ is similar to that of inversion, but with an important difference. In the case of inversion, the Bloch states only have well-defined parities at the eight TRIM, which are the reciprocal-space points that are carried onto themselves by inversion. By the same token, the parity analysis of the HW states discussed above is only applicable at the four PTRIM. As a result, Dirac touchings between Wannier bands may occur at $z = 0$ or $z = c/2$ at the four PTRIM, but are generically absent elsewhere unless enforced by additional symmetries.

The $M_z$ operation, on the other hand, acts as the identity on $k_x$ and $k_y$, so that all $k$ in the 2DBZ follow “mirror parity” rules similar to those discussed above for the PTRIM under inversion. That is, at any $k$, the Wannier bands can be decomposed into those centered at $z = 0$ with even or odd parity labels; those centered at $z = c/2$ with even or odd parity labels about their center; and mirror-symmetric pairs of Wannier bands at $\pm z$. A consequence is that Wannier bands can be pinned exactly at $z = 0$ or $z = c/2$ over the entire 2DBZ.

Moreover, in the presence of mirror symmetry, a new kind of topological invariant known as the “mirror Chern number” comes into play [3, 51, 113]. The mirror Chern number on the $k_z = 0$ symmetry plane is normally defined in reciprocal space in terms of the difference of Chern indices of the even- and odd-parity mirror subspaces on this plane, whose points are left invariant under $M_z$. Except for certain centered lattices [125], a second mirror Chern number can be defined in the same way on the $k_z = \pi/c$ plane as well.

In a separate collaboration involving some of us, we have investigated the Wannier bands in the presence of $M_z$ symmetry in some detail, clarifying the generic behaviors that are expected, and discussing the rules for deducing not only the axion $Z_2$ index, but also the mirror Chern number, from the Wannier band structure. The work will be published elsewhere [96].

*Alternating Haldane model* — To illustrate the case of mirror symmetry we use a spinless tight-binding model introduced in Olsen et al. [86], which we refer to as the “alternating Haldane model.” This model consists of layers of the Haldane model [47] stacked along the (001) direction, with two layers per cell. The Hamiltonian can be
expressed as

\[ H = \sum_p \left[ H_p + t_{3,p} \sum_i \tau_i \left( c_{p,i}^\dagger c_{p+1,i} + \text{h.c.} \right) \right] \] (4.41)

where the first term describes isolated layers, the second term couples adjacent layers, and “h.c.” stands for “Hermitian conjugate.” The Hamiltonian is constructed in such a way that in the decoupled limit adjacent layers have opposite Chern numbers. Specifically,

\[ H_p = (-1)^p \Delta \sum_i \tau_i c_{p,i}^\dagger c_{p,i} + t_1 \sum_{(ij)} c_{p,i}^\dagger c_{p,j} \]
\[ + (-1)^p t_2 \sum_{\langle ij \rangle} i \nu_{ij} c_{p,i}^\dagger c_{p,j} \] (4.42)

where indices \( i \) and \( j \) label the honeycomb sites and \( \tau_i = +1(-1) \) for \( i \in A(B) \), where \( A \) and \( B \) are the two honeycomb sublattices. The first and second terms contain the on-site energies and nearest-neighbor hoppings respectively, and the third describes a pattern of staggered magnetic fluxes generated by complex second-neighbor hoppings. Therein, \( \nu_{ij} = +1 (-1) \) if the hopping direction from \( j \) to \( i \) is right-handed (left-handed) around the center of a plaquette [47]. The \((-1)^p\) factor in the third term is responsible for reversing the Chern number between adjacent layers. The first term also contains a \((-1)^p\) factor that reverses the on-site energies; as a result, adjacent layers are not simply time-reversed as in the case of “antiferromagnetic TIs” with \( \{E'|c/2\} \) symmetry [79].

The above model was used in Olsen et al. [86] to pump a quantum of axion coupling during a slow cyclic evolution of the Hamiltonian along the path

\[ t_1 = 4.0, \]
\[ t_2 = -1.0, \]
\[ t_{3,p} = -(1 + (-1)^p - 1) \sin \phi, \]
\[ \Delta = -(3\sqrt{3} + 2 \cos \phi), \] (4.43)

parameterized by the angle \( \phi \). Here we are interested in the configurations \( \phi = 0, \pi \) where \( t_{3,p} \) becomes independent of \( p \), and as a result the model acquires mirror symmetry about the planes of the layers. At \( \phi = 0 \) the half-filled system is axion-even, and at \( \phi = \pi \) it is axion-odd [86].
Figure 4.8: Wannier band structures for the two mirror-symmetric phases of the alternating Haldane model parametrized by the angle $\phi$ in Eq. (4.43). (a,c) is the trivial-insulator phase with $\phi = 0$, and (b,d) is the axion-insulator phase with $\phi = \pi$. The wannierization direction is (001) – normal to the mirror plane – in (a,b), and (010) (denoted $\hat{z}'$) – lying in the mirror plane – in (c,d). In (a,b), we indicate the Chern numbers of the Wannier bands to emphasize that while the Wannier band structures look identical, they actually describe different phases.

We first wannierize the axion-even and odd systems along the direction (001) normal to the mirror plane, as done in Appendix A of Olsen et al. [86]. Within this setting the mirror operation becomes $M_z$, and the resulting Wannier bands are shown in the top panels of Fig. 4.8. The earlier discussion indicates that a consequence of $M_z$ symmetry is that entire bands can be pinned exactly at $z=0$ or $z=c/2$ over the full 2DBZ. As is clear from the figure, this is the case here. In fact, with just two occupied bands, our model is so simple that this is all we have, with one Wannier band pinned at $z=0$ and the other at $z=c/2$. Just as for the inversion-symmetric case, the axion angle $\theta_{cs}$ is 0 or $\pi$ depending on whether the Chern index of the boundary-centered group is even or odd. This is confirmed by inspection of Fig. 4.9, where we show the Chern numbers of
the Wannier bands in a slab geometry.

Finally, we consider the case of wannierization along an axis lying in the mirror plane, specifically, along (010), which we denote as $z'$. Now the mirror operation is of type $M_d$, and the calculated bulk Wannier bands are displayed in the bottom panels of Fig. 4.8. Figure 4.8(c) shows a disconnected band structure, and consulting Sec. 4.2.2 we conclude that the system is axion-trivial. Conversely, in the axion-odd phase the Wannier bands are now required to flow, each with an odd number $N_d$ of down-touchings. This is indeed what happens in Fig. 4.8(d), where $N_d = 1$ for all bands.

Figure 4.9: Chern index of each isolated slab Wannier band of the alternating Haldane model, plotted as a function of its average position $\bar{z}$. We consider slabs with a thickness of 10 unit cells along (001) in (a) the trivial phase of Fig. 4.8(a), and (b) the axion phase of Fig. 4.8(b). In the trivial phase each Wannier band carries a zero Chern index, while in the axion-odd phase the Chern index alternates between $-1$ and $+1$.

4.4.4 Glide mirror symmetry

General considerations — A $\mathbb{Z}_2$ classification of 3D insulators with glide mirror symmetry was introduced in Fang and Fu [37] and Shiozaki et al. [108], and was later argued to be equivalent to the axion $\mathbb{Z}_2$ classification [125]. Interestingly, glide mirror symmetry realizes all three types of axion-odd space-group symmetries, depending on the choice of wannierization direction. If $\hat{z}$ is chosen as the normal direction to the reflection plane, the glide operation becomes $\{M_z|\tau_\perp\}$; recalling that our classification discards in-plane fractional translations $\tau_\perp$, here the quantizing symmetry is simply
$M_z$, a $z$-reversing operation, as in Sec. 4.2.2. Instead, we can choose $\hat{z}$ to lie in the reflection plane and be perpendicular to the half-lattice translation. In this frame the glide operation is $\{M_d|\tau_\perp\}$; again discarding $\tau_\perp$, the axion-odd operation becomes $M_d$, a symmorphic $z$-preserving operation, per Sec. 4.2.2. Finally, if $\hat{z}$ is chosen along the half-lattice translation, then the glide operation becomes $\{M_d|c/2\}$, a nonsymmorphic $z$-preserving operation.

In the numerical tests below we shall consider the latter nonsymmorphic scenario, which we have not encountered in our previous examples, as well as the $z$-reversing case. We will omit the symmorphic $z$-preserving configuration, since we have already encountered this kind of $M_d$ symmetry in our study of the alternating Haldane model.

Model — As an example, we take the four-band tight-binding model at half filling described in Fang and Fu [37] (an equivalent model was also used in Varjas et al. [125]). The simple orthorhombic $a \times b \times c$ unit cell contains two pairs of orbitals with reduced coordinates $(0,0,0)$ and $(0.5,0,0)$, and the Hamiltonian is expressed in $k$ space as

$$H_k = \sin \left(\frac{k_x a + \phi}{2}\right) \rho_x \tau_x + \sin(k_y b) \rho_0 \tau_y + \sin(k_z c) \rho_z \tau_z + (m - \cos k_x a - \cos k_y b - \cos k_z c) \rho_0 \tau_z.$$ (4.44)

The Pauli matrices $\tau$ can be considered to represent either spin or orbital degrees of freedom, and the matrices $\rho$ are associated with sublattice degrees of freedom.

The above model is invariant under the glide operation $\{M_z|a/2\}$ consisting of a reflection about the $z=0$ plane followed by a half-lattice translation along $\hat{x}$. Because it is so simple, the model has two additional axion-odd symmetries, $\{M_y|a/2\}$ and $M_x$, whose presence obscures the role of $\{M_z|a/2\}$ in quantizing $\theta_{cs}$. For example, the mirror $M_x$ forces the $(100)$-oriented Wannier bands to be completely flat. To break these extra symmetries, we displace the orbitals to $(0,0,0.1)$ and $(0.5,0,-0.1)$, and include an additional term

$$V_k = 0.4 \cos \frac{k_x a}{2} \rho_x \tau_z$$ (4.45)

in the Hamiltonian, chosen weak enough so as not to cause any band inversion. At half filling, the two lowest bands are filled. The model has two adjustable parameters; we set $\phi = 0.4$, and vary $m$ to access different phases.
Figure 4.10: Wannier band structures for the model with glide mirror symmetry. The trivial and axion phases were obtained by setting $(m, \phi) = (3.5, 0.4)$ and $(m, \phi) = (2.0, 0.4)$ respectively, in Eq. (4.44). In (a,b) the wannierization direction is $\hat{x} \parallel (100)$, lying in the glide plane and along the half-translation; in (c,d) it is $\hat{z} \parallel (001)$, normal to the glide plane. The Chern numbers of isolated bands are indicated in panels (a-c).

Figure 4.10 shows the Wannier band structures calculated with $m = 3.5$ (trivial phase), and with $m = 2.0$ (axion phase), for two different wannierization directions. In the upper panels the wannierization is along the half-lattice translation (along $\hat{x}$), and in the lower panels it is normal to the reflection plane (along $\hat{z}$). (For this model, instead of reorienting the axes so that the wannierization direction is always along $z$ as in the previous examples, we have chosen to keep the axes fixed.)

In the upper panels of Fig. 4.10, the glide operation is $\{M_d|a/2\}$ in the notation of Table 4.1. This is a nonsymmorphic operation that preserves the wannierization direction, of the type discussed in Sec. 4.2.2 for the nonsymmorphic operation preserving $\hat{z}$ case. In axion-odd phases protected by such symmetries, the Wannier spectrum is not required to flow. (This is in contrast to axion-odd phases protected by a simple mirror $M_d$, which must exhibit flow as illustrated in Fig. 4.8(d) for the alternating Haldane
model.) Indeed, the spectrum of the present model is gapped not only in the trivial phase of Fig. 4.10(a), but also in the axion-odd phase of Fig. 4.10(b). We can therefore use Eq. (4.34) to determine the axion index, taking for $C_{LG}$ the Chern number of either band. The Chern numbers of the two bands are $(0,0)$ in Fig. 4.10(a) and $(-1,+1)$ in Fig. 4.10(b), leading to $\theta_{cs} = 0$ and $\theta_{cs} = \pi$ respectively.

Consider now the lower panels of Fig. 4.10, where the glide operation returns to $\{M_z|a/2\}$ in the notation of Table 4.1. This is a $z$-reversing operation, like spatial inversion or a simple mirror $M_z$, so we refer to Sec. 4.2.2 for the $\mathbf{z}$-reversing case. Now the Wannier spectrum is gapped in the trivial phase of Fig. 4.10(c), but connected in the axion phase of Fig. 4.10(d). (This is different from the alternating Haldane model with $M_z$ symmetry, which exhibited gapped Wannier spectra in both phases.) In the trivial phase of Fig. 4.10(c), the two isolated bands have vanishing Chern numbers, so that Eq. (4.29) gives $\theta_{cs} = 0$. For the axion phase of Fig. 4.10(d), which exhibits “fragile” flow, we resort to Eq. (4.31), and since there are single Dirac cones at integer and half-integer values of $z_n/c$, we conclude that $\theta_{cs} = \pi$. At variance with our previous examples of connected band structures, here the Dirac nodes are located at generic low-symmetry points in the 2DBZ, not at high-symmetry points or lines.
Chapter 5

Controllable quantum point junction on the surface of an antiferromagnetic topological insulator

5.1 Introduction

In this work, we develop a theoretical prescription for the creation and manipulation of chiral edge channels on the surface of an antiferromagnetic (AFM) TI. This class of materials was introduced theoretically by Mong and Moore [79] and has recently become the focus of intense research with various candidates such as MnBi$_2$Te$_4$ [87], MnBi$_4$Te$_7$ [52], EuIn$_2$As$_2$ [144] and NpBi [145] appearing in the literature. Motivated by these recent developments and the fact that there is in principle no reason why both the bulk and surface gaps could not be on the order of hundreds of meV, allowing for potential high temperature device operation for certain applications, we propose and explore the properties of a robust and controllable quantum point junction (QPJ) on the surface of an AFM TI.

Figure 5.1(a,b) shows a prototypical spin arrangement in an AFM TI. The magnetic ordering is A-type AFM, i.e., with magnetization uniform in-plane but alternating from plane to plane along the stacking direction, which we take to be along $\hat{z}$. As described in Mong et al. [79], each individual layer can be thought of as adiabatically connected to a 2D Chern insulator, with the sign of the Chern number alternating from layer to layer. The sign of the surface anomalous Hall conductivity of $\pm e^2/2h$ is thus determined by the magnetic orientation of the last layer at the surface. As a result, two kinds of 1D chiral channels can occur at the surface. As shown in Fig. 5.1(a), the emergence of a bulk AFM domain wall at the surface reverses the sign of the anomalous Hall conductivity on either side of the resulting line defect, which therefore carries a topologically protected chiral channel we refer to as a domain-wall channel. Alternatively, even if no bulk AFM
domain walls are present, a single-height step can occur on the surface, as shown in Fig. 5.1(b). If it does, it also marks a sign reversal of the anomalous Hall conductivity when crossing the step, and thus carries a chiral edge channel as well. We will refer to this as a step channel.

Figure 5.1(c,d) shows the manifestation of the domain-wall and step channel in the surface band structure as described in the context of a tight-binding model used throughout this work (see Methods). The presence of either of these defects results in 1D linear dispersions in the otherwise gapped bulk and surface spectrum of the AFM TI. The states that comprise the chiral bands are exponentially localized in the vicinity of the channel, and host 1D massless Dirac fermions.

Figure 5.1: Types of chiral channels at surface of an A-type antiferromagnetic topological insulator (AFM TI). Depiction of the chiral channel (blue cylinder) due to (a), a bulk domain wall (b), a surface step. Surface band structures along (001) in the presence of (c), a bulk domain wall (d), a surface step. The projection of the states on the chiral channels (blue cylinder) in (a,b) are also shown (blue markers) to illustrate the localization of the massless Dirac fermions that disperse linearly along the channel direction at low energy with velocities (c), $v_{dw}$ and (d), $v_{st}$. The description of the model Hamiltonian can be found in Methods. Energies are expressed in terms of the onsite energy $m$ in Equation (5.11).
The novel opportunity opened by the presence of two different kinds of 1D chiral channels at the surface is that these can be made to intersect, as shown in Fig. 5.2(a), and such intersections are expected to remain thermodynamically stable. In contrast, as illustrated in Fig. 5.2(b), an intersection between two surface steps can easily evolve via a pinch-off event into a configuration in which an isthmus of constant surface height separates the steps; indeed, the width of such an isthmus will tend to grow due to the line tensions of the steps, and the quantum junction will be removed. A similar mechanism affects the intersection of two domain walls [97]. In fact, setups like those depicted in the inset of Fig. 5.2(b), where two chiral channels come in close proximity, have long been used in quantum Hall systems to realize electron interferometers [28, 57]. These constructions, known as quantum point contacts (QPCs) [123, 138], enable tunnelling between channels, and were recently used to observe the braiding of anyons [82].

Our proposal aims to highlight a robust way to construct intersecting chiral channels by making use of a novel material system that is on the verge of discovery. In fact, these junctions were recently observed to appear naturally at the surface of the putative AFM TI MnBi$_2$Te$_4$ [105]. Moreover, we show that a QPJ can be controlled by scanning tips of the kind used in scanning tunneling microscopy (STM) and related methods. Here, we are interested in local probes that affect the magnetic moments and electrostatic potential, which we refer to as magnetic and electrostatic STM tips respectively. We explore the properties of the QPJ by constructing the Hamiltonian associated with the system depicted in Fig. 5.2(a) and performing dynamic wave-packet (WP) simulations that allow us to extract the $S$-matrix of the junction. Remarkably, we find that magnetic and electrostatic STM tips in proximity with the junction can, in principle, be used to realize any unitary $S$-matrix. In addition, we show that the effect of symmetry breaking terms and weak disorder can be “gauged away” using the two tips.

The stability and tunability of the proposed junction, together with the intrinsic benefits of a magnetic topological material, can be of great utility in established and new applications involving unidirectional channels, such as electron interferometry, low-power electronics and quantum information processing.
Figure 5.2: Stable and unstable junctions (a), The intersection of a domain-wall channel with a step channel results in a thermodynamically stable junction, i.e., small surface deformations can only move the junction but not remove it. (b), The intersection of two step channels (or two domain-wall channels) is unstable. The inset shows how small deformations remove the junction. Blue arrows indicate the direction of propagation on the chiral channels, while orange and yellow surfaces indicate whether the anomalous Hall conductivity is $\pm e^2/2h$ respectively.

5.2 Results

5.2.1 Extracting the $S$ matrix

We begin by considering the WP dynamics at the surface of an AFM TI. Figure 5.3(a), shows the calculated time evolution of a WP on a single domain-wall channel, while Fig. 5.3(b), that of a WP in the presence of the QPJ in Fig. 5.2(a). In both cases the dissipationless channels are protected from back-scattering by the insulating bulk and surface gaps. The wave function of the WP is thus exponentially confined to the vicinity of the one-dimensional channel, and it travels with a constant group velocity along the channel. In Fig. 5.3(b), a WP enters along the domain-wall channel, gets split by the QPJ, and then the two components travel away from the QPJ along the step channels. Later we shall consider configurations in which multiple consecutive scattering events occur.

To understand the behaviors observed above, we note that the wave function of a WP propagating along a single domain-wall channel in direction $y$, as in Fig. 5.3(a),
can be well approximated as

\[ \Psi(x, y, z, t) = \chi(x, z) f(y - y_0 - v dt). \] (5.1)

Here \( \chi(x, z) \) captures the transverse shape \((x, z)\) and spin-orbital character \((\sigma, \tau)\) indices respectively) of the WP, while \( f(y) \) is the envelope function of the WP, which we take to be a Gaussian. The WP is launched from position \( y_0 \) at time \( t = 0 \) and travels with group velocity \( v_{\text{dw}} \) (which is set by the surface state dispersion in Fig. 5.1(c)). In modeling at this level we neglect spreading of the WP, which we find to be negligible in our simulations. Similar considerations apply to the propagation of a WP on a step along \( x \) traveling with group velocity \( v_{\text{st}} \) (that is set by the surface state dispersion in Fig. 5.1(d)).
We now consider the scattering event depicted in Fig. 5.3(b), where an incoming WP splits after encountering the QPJ. We will use unprimed labels $a$ and $b$ to refer to the two incoming domain-wall channels of Junction 1, as in Fig. 5.3(c). Note that the extra junctions are the result of in-plane periodic boundary conditions. The incoming initial conditions are specified by amplitudes $\phi_a = 1$ and $\phi_b = 0$. Now let $t_1$ indicate a time after the scattering through Junction 1 is complete, but before Junction 2 is encountered. We label the two outgoing step channels as $a'$ and $b'$, adopting once and for all the arbitrary convention that $a \rightarrow a'$ and $b \rightarrow b'$ result from taking left turns, as shown in Fig. 5.3(c). As illustrated in Fig. 5.3(b), one component of the WP moves to the right and the other to the left, with velocities $v_{st}$ and $-v_{st}$ respectively. At time $t_1$ both will be centered at a distance $x_1$ relative to the junction, so in general we expect to find

$$\Psi_{\sigma r}^{st}(x, y, z, t_1) = \phi_a' \tilde{\chi}_{\sigma r}^{st}(y, z)f(x + x_1) + \phi_b' \chi_{\sigma r}^{st}(y, z)f(x - x_1).$$ (5.2)

Here $\phi_a'$ and $\phi_b'$ are the amplitudes (magnitude and phase) describing scattering from incoming channel $a$ into channels $a'$ and $b'$ respectively, and $\tilde{\chi}_{\sigma r}^{st}$ is the time-reversed partner of $\chi_{\sigma r}^{st}$. These expectations are well reproduced in our full numerical calculations which therefore allow us to extract the amplitudes $\phi_a'$ and $\phi_b'$.

Similar calculations, where the incident WP approaches Junction 1 along the $-\hat{y}$ direction on channel $b$, allow us to extract the corresponding amplitudes that result for initial conditions of $\phi_a = 0$ and $\phi_b = 1$. Thus, we can model a combined scattering event via

$$\begin{pmatrix} \phi_{a'} \\ \phi_{b'} \end{pmatrix} = S \begin{pmatrix} \phi_a \\ \phi_b \end{pmatrix}$$ (5.3)

where the elements of the $S$-matrix are determined by the four complex amplitudes discussed above.

In this way, the evolution of the system of propagating WPs is mapped onto that of a two-level quantum system, so that it is enough to restrict $S$ to be an SU(2) matrix.\footnote{More generally, systems that involve many junctions and result in $N > 2$ output channels will...}
The characterization of a junction by such an $S$-matrix is a central element of our theory. It is illustrative to represent the initial or final state as a point on the Bloch sphere,

\[
\begin{pmatrix}
\phi_a \\
\phi_b
\end{pmatrix} = \begin{pmatrix}
\cos(\theta/2) \\
e^{i\phi} \sin(\theta/2)
\end{pmatrix},
\]

(5.4)

where $\theta$ determines the relative WP magnitude on channels $a$ and $b$ and $\phi$ their phase difference, as illustrated in Fig. 5.3(d).

Each junction scattering event can then be described by the action of the corresponding junction $S$-matrix on the spinor representation of the channel states, regarded as a qubit state, and the result of consecutive QPJ scattering events, as in Fig. 5.3(e), corresponds to the action of consecutive gates acting on these qubits as illustrated in Fig. 5.3(f).

Let us now return to a more specific discussion of our full time-evolution calculations, and our analysis of them in terms of the framework sketched above. Figure 5.3(e), shows the time evolution of a WP initiated on channel $a$. The WP propagates towards and then scatters at Junction 1, splitting into two equal parts. Later the two WPs pass through Junction 2, interfering destructively and constructively on outgoing channels $a'$ and $b'$ respectively. As promised, we can describe the time evolution of the WP configuration as a qubit passing through two gates. Indeed, using the convention of Fig. 5.3(c), the calculated $S$ matrix of Junction 1 and 2 corresponds to the Hadamard gate

\[
S_1 = S_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix},
\]

(5.5)

so that the final state is related to the initial one by applying the Hadamard gate twice. Geometrically the $S$ matrix expressed as

\[
S = R_\mathbf{n}(\varphi) = e^{-i\mathbf{n}\cdot\sigma/2}
\]

(5.6)

describes a qubit rotation by an angle $\varphi$ through an axis $\mathbf{n}$ and $\sigma = (\sigma_x, \sigma_y, \sigma_z)$ is a vector of Pauli matrices. Since $S_1 = S_2 = R_\mathbf{y}(\pi/2)$, each application rotates the qubit transform under $S \in \text{SU}(N)$. In such cases each junction acts as $S \in U(2) \subset \text{SU}(N)$. In this work we are only concerned with systems with $N = 2$, so that each junction acts as an $\text{SU}(2)$ matrix.
Figure 5.4: Magnitude and phase manipulation of the quantum point junction. (a), The magnetic tip with $V_Z = 2mZ$ in Equation (5.14) (see Methods), has polarized the surface spins in a circular region centered at the junction resulting in two uncoupled channels. (b), Partially polarized region with $V_Z = 0.4mZ$ causes unequal splitting of the wave packet (WP). (c), Numerical calculation of the magnitude splitting $\cos^2 \beta/2$ as a function of $V_Z$. Each value corresponds to the integral of $|\Psi(t_1)|^2$ on channel $a'$ of Junction 1. (d),(e), Applying the electrostatic tip with $V_G = 0.6$ in Equation (5.14), at Junction 1 induces a phase difference between the outgoing WPs which then affects how they interfere at Junction 2. (d), The center of the rectangular region $\Omega_G$ is chosen so that $\gamma = \pi/2$. (e), Same but we set $\gamma = \pi$ making the WPs constructively (destructively) interfere on channel $a(b)$ of Junction 2. (f), Numerical calculation of the angle $\gamma$ from the relative phase of the outgoing WPs at $t_1$ as a function of $x_0$. The phases of the outgoing WPs are determined from the inner product between $\Psi(t_1)$ in the presence and absence of the phase gate.

by $90^\circ$ around the $\hat{y}$ axis of the Bloch sphere, resulting in an overall reversal of the pseudospin as shown in Fig. 5.3(f).
### 5.2.2 Controlling the $S$ matrix

Before explaining how the control of the $S$ matrix is achieved, it is illustrative to break down the action of $S$ into three stages. First we have the propagation along the incoming channels; since these cannot scatter into one another, this is represented by a diagonal matrix $S_{\text{dw}}$. Then there is the scattering $S_{\text{pj}}$ at the QPJ itself, followed by another channel-diagonal propagation $S_{\text{st}}$ on the outgoing step channels. The overall $S$ matrix can then be written in terms of the Pauli matrices as

$$S = S_{\text{st}}S_{\text{pj}}S_{\text{dw}} = e^{-i\gamma\sigma_z/2}e^{-i\beta\sigma_y/2}e^{-i\alpha\sigma_z/2},$$

(5.7)

where $S_{\text{pj}}$ is expressed as a real orthogonal matrix because the phases can be absorbed into $S_{\text{dw}}$ and $S_{\text{st}}$. Remarkably, $(\alpha, \beta, \gamma)$ are exactly the three Euler angles that can be used to express any SU(2) matrix. Thus control over the three Euler angles results in a universally programmable gate, which we now demonstrate.

To control the $S$ matrix, we will use two local probes in the vicinity of the junction. The first one, which we refer to as a magnetic STM tip, affects the local magnetic moments, and as we shall see, controls the magnitudes of the $S$ matrix. The second probe is an electrostatic STM tip modifying the site energies under the tip, thus controlling the phases of the $S$ matrix. The effect of the magnetic tip is controlled through the coefficient $V_Z$, and that of the electrostatic tip through $V_G$, both acting in the local vicinity of the junction. For more details on how this is modeled, see Equation (5.14) in Methods.

**Magnitude control.** — We set $V_G = 0$, leaving the electric potential constant throughout the crystal so that no extra phase evolution occurs during the propagation ($\alpha = \gamma = 0$), and we vary the strength of the magnetic tip $V_Z$. This affects the left-right magnitude splitting, i.e., $S = R_\beta = e^{-i\beta\sigma_y/2}$ in Equation (5.7), with $\beta = \beta(V_Z)$. To understand the mechanism behind the magnitude control, first consider the extreme scenario depicted in Fig. 5.4(a). Here a strong magnetic STM tip has polarized the surface magnetization in the vicinity of the junction (orange circular region), forcing the anomalous Hall conductivity to be uniformly $+e^2/2h$ in that area (see Methods). This essentially “removes” the junction, and the WP is completely transferred from the
domain-wall (channel $a$) to the edge of the step (channel $b'$), so that $S_1 = R_y(\pi)$. An example of partial polarization, is shown in Fig. 5.4(b), while the results of tuning $V_Z$ over the entire range of tip strength is shown in Fig. 5.4(c), where we plot the numerically calculated value of $\cos^2(\beta/2)$, which represents the asymmetry between left- and right-scattered WPs, as a function of $V_Z$. This demonstrates the universal control of the Euler angle $\beta$ using a magnetic STM tip. In the Sec. 5.2.3 we consider a strong magnetic STM tip that decouples the channels inducing a QPC and we vary the area of the region applied to analyze tunneling between the channels.

**Phase control.** — To illustrate the phase control, we set $V_Z = 0$, fix $V_G$ to a non-zero value (see Equation (5.14) in Methods), and control the position of the electrostatic tip. Then $S_j = R_z(\gamma)R_y(\pi/2)R_z(\alpha)$, where $\alpha$ and $\gamma$ are determined by the position $(x_0, y_0)$ of the tip relative to the junction, as described by Equation (5.15) (see Methods). The electrostatic tip is depicted as a shaded square with origin $(x_0, y_0)$ in Fig. 5.4(d),(e). In fact, our choice of $\phi_a = 1$ and $\phi_b = 0$ simplifies the situation, since $R_z(\alpha)$ just corresponds to an overall phase, which is not of interest. Physically, the WP splits equally at the first junction ($\beta = \pi/2$), and the electrostatic STM tip, corresponding to the second term in Equation (5.14), is then used to control the relative phases of the outgoing WPs via the $R_z(\gamma)$ term.

In Fig. 5.4(d)-(f), we illustrate the phase control by applying the electrostatic gate on Junction 1. To see the effects of the phase manipulation, we consider the interference that conveniently occurs when the WPs meet again (due to periodic boundary conditions in $x$ and $y$) at Junction 2. In Fig. 5.4(d), the electrostatic tip is centered four unit cells to the right at $(x_0, y_0) = (4, 0)$, which approximately makes $\gamma = \pi/2$ so that $S_1 = R_z(\pi/2)R_y(\pi/2)$, while $S_2 = R_y(\pi/2)$ as before. After scattering at Junction 1 the outgoing WP, whose state corresponds to a vector pointing along the $+y$ direction of the Bloch sphere, becomes the incoming WP at Junction 2. Since Junction 2 acts as a rotation around the $y$-axis, it does not affect the qubit state of the WP. Similarly, in Fig. 5.4(e), we set $(x_0, y_0) = (8, 0)$, so that after encountering Junction 1 the qubit state points along $-\hat{x}$, and after Junction 2 it returns to its initial $+\hat{z}$ state. In Fig. 5.4(f), we present a numerical calculation of $\gamma$ versus $x_0$. This is done by calculating the phase
of the WPs just after it scatters off Junction 1. We find a linear behavior as expected from Equation (5.15).

In summary, using the two STM tips we can control $\alpha$, $\beta$, and $\gamma$ independently in Equation (5.7), so that the junction can be made to implement any SU(2) gate.

5.2.3 Robustness of junction

Channel tunneling — In the Sec. 5.2 we demonstrated the control of magnitude splitting by applying the magnetic tip in a circular region (of constant radius) centered at the junction and varying the strength $V_Z$ of the tip (see Eq. 9). Here instead we apply a magnetic tip with constant $V_Z = 2mZ$, i.e., strong enough to polarize the spins and decouple the channels, and we vary the radius of the region applied. In this way, we are effectively modeling a quantum point contact since a WP can scatter to the decoupled channel only through tunneling. Furthermore, this approach enables the extraction of the tunneling length $\xi$.

In Fig. 5.5(a), we plot the transmission $T = \cos^2 \beta/2$, i.e., the magnitude of the wave function that has remained on the channel, to show the exponential suppression of tunneling to the decoupled channel. Since, $T(r = 0) = 0.5, T(r \to \infty) = 1.0$ we heuristically expect $T \simeq 1 - e^{-2r/\xi}/2$ with the 2 in exponent coming from the fact that $T$ is related to the square of the wavefunction. In Fig. 5.5(b), we replot the data on rescaled axes to extract the tunneling length, which is approximately $\xi = 0.48\alpha$.

---

Figure 5.5: Tunneling between decoupled channels. (a), Transmission as a function of the radius of the polarized region. The radius is measure in units of the lattice constant $\alpha$. (b), Linearizing the plot in (a), to extract the tunneling length $\xi$. (c), Example of tunneling for polarized region with $r = 3\alpha$. 
Breaking symmetries — The simple model Hamiltonian we are using has many symmetries. These extra symmetries do not affect any of the core features of our proposal such as the unidirectionality and topological protection of the channels or the robustness and controllability of the junction. However, they result in some features that should not be generically expected in any realistic implementation. In this section we add symmetry breaking terms to the Hamiltonian and illustrate how these non-essential features are affected. In particular, we find that mirror-symmetry-breaking terms affect the amplitude splitting and particle-hole-symmetry-breaking terms the group velocities of the domain-wall and step channels.

Figure 5.6: Effect of symmetry breaking. (a-c), A mirror-symmetry-breaking term modifies the magnitude splitting at the junction (d-f), A particle-hole-symmetry-breaking term enhances the group-velocity anisotropy of the domain-wall and step channels. (a), Magnitude \( \cos^2(\beta/2) \) and phase \( \gamma \) as a function of the strength \( t_M \) of the mirror-breaking hopping term \( V_M \). Only the magnitude is modified by the addition of this term. (b), Setting \( v = 0.24 \) in Equation (5.8) causes unequal splitting of the WP but we can use the magnetic tip with \( V_Z = 0.14 \) (c), to cancel the effect of the mirror-breaking term. (d), Group velocity of the domain-wall and step channel as a function of the strength \( t_{PH} \) of mirror-breaking term \( V_{PH} \). Time-evolution without (e), and with (f), the inclusion of \( V_{PH} \). The effect is to slow down the WP but has no effect on the S-matrix.

Mirror symmetry — As discussed in the main text, the fact that the QPJ “naturally”
implements the Hadamard gate is a manifestation of the high symmetry of the model
and of the geometry of the junction. To this effect we break the extraneous $M_x$ and
$M_y$ mirror symmetries and study how they affect the S-matrix. To do that we add the
term
\[ V_M = \frac{-i t_M}{2} \sum_{\ell \ell'} c_{\ell}^\dagger x^\prime \sigma^x c_{\ell'} \delta_{\ell_z \ell'_z}, \]
(5.8)
where $\delta_{\ell_z \ell'_z}$ ensures that only in-plane hopping terms are considered. Note that this
term couples spin and orbit in a way that preserves $T$ symmetry [139]. Even though
Equation (5.8) does not affect the magnetic properties of the junction it nonetheless
affects the S-matrix by modifying the magnitude splitting $\cos^2(\beta/2)$, as the numerical
calculations indicate in Fig. 5.6(a). Importantly, the universal control over the S-matrix
magnitudes using the local magnetic tip enables us to eliminate (i.e., calibrate) the effect
of the mirror-symmetry-breaking terms, as illustrated in Fig. 5.6(b).

Particle-hole symmetry — A secondary feature of the model is that the group ve-
locities of the domain-wall and step channels are approximately equal $v_{dw} \approx v_{st} = v$
(less than 1% difference). We can enhance the anisotropy by including a particle-hole-
symmetry-breaking term
\[ V_{PH} = \frac{t_{PH}}{2} \sum_{\ell \ell'} c_{\ell}^\dagger c_{\ell'} \delta_{\ell_x \ell'_x} \delta_{\ell_y \ell'_y}, \]
(5.9)
where again we make use of delta functions to ensure only hoppings along specific
directions are considered, in this case only along the $z$-direction. Fig. 5.6(d), shows that
the effect of $V_{PH}$ is to reduce the group velocities of both channels in an inequivalent
way.

It is important to realize that for any junction, both of the incoming (or outgoing)
channels are of the same type, either both domain-wall or both step. Thus, even if the
velocities of the two channels are unequal, this will not cause a relative delay between
the arrival of the two wave packets at the junction which would affect their interference.
This is also confirmed by the explicit time-evolution simulations in Fig. 5.6(e,f), where
we illustrate how the addition of $V_{PH}$ slows down the WPs but does not effect any of
our conclusions.

Stability to disorder — Finally we show that the QPJ is robust in the presence
Figure 5.7: Effect of disorder. (a,b), Averaged magnitude splitting $\cos^2(\theta/2)$ and phase $\gamma$ over 20 disordered configurations for different values of $W$. (a), Because the initial WP $\Psi(0)$ is not an exact eigenstate there is loss to the bulk we need to take into account. The green line shows the magnitude of the WP that remained on the surface after scattering at Junction 1 and the red line corresponds to the renormalized magnitude splitting. The disorder affects slightly the magnitude splitting which fluctuates about 50% for different realizations of disorder. (b), The phase $\gamma$ is significantly affected by disorder and needs to be accounted. (c), Example of a specific disorder realization as a function of $W$. (d), For the realization in (c), with $W = 0.1$ we can use the gate voltage tip to eliminate the random phase difference accumulated up to $t = t_2$, due to the disorder. In this example we set the strength of the gate voltage to $V_G = 0.1$ and the center of the rectangle at $(x_0, y_0) = (8, 0)$.

We introduce disorder into the model Hamiltonian by adding a random potential scattering term given by

$$V_D = \sum_{\ell} m_D(\ell) c_{\ell}^\dagger c_{\ell}. \quad (5.10)$$

The disorder potential $m_D(\ell)$ is sampled from a Gaussian distribution at each site in the three-dimensional lattice $\ell$ with zero mean and standard deviation $W$, which characterizes the strength of the disorder potential. For sufficiently large $W$ the average band gap in the bulk will close and the model will transition out of a topological phase. In contrast, for weaker disorder strengths (relative to the clean band gap, which is equal
to $E_g \approx 0.6$ here) the topological properties are expected to remain robust. This should also provide a level of protection of the chiral surface states from back-scattering due to the disorder potential, but how this impacts the quantum point junction remains unclear.

To demonstrate the robustness of the junction to disorder, we have performed a similar analysis as in the previous section. Of course, we cannot construct the initial WP from the exact eigenstates because of the presence of disorder in the system. For that reason, we use the same initial WP that was constructed for the clean model, and renormalize our results by accounting for the prompt loss of amplitude associated with the fact that this trial function has nonzero overlap with some extended bulk and surface states (see Fig. 5.7(a)). The results are averaged over 20 disorder samples and show that magnitude and phase evolution through the junction are randomly effected by disorder, and therefore retain the same average value as they do in the absence of disorder. In particular, the random potential induces both a splitting of the WP magnitude as well as a shift in the phase that are both random for each sample, which clearly averages to zero as demonstrated in Fig. 5.7(a,b).

Lastly, we turn to how disorder can impact the QPJ in a specific disorder realization. We show that the effect of disorder can be calibrated adopting an approach like that used above for the mirror-breaking term. As shown in Fig. 5.7(c), for a specific random sample, the disorder does not affect the splitting, at least for small $W$, while the phase of the WP is significantly different from the clean limit. We then apply the gate voltage to completely remove this effect, returning the QPJ to its clean behavior, Fig. 5.7(d). These results demonstrate both a robustness and a level of control over disorder in this novel QPJ. This is important, as each device made out of such a QPJ will have some random disorder profile. Nevertheless, as we have shown in Fig. 5.7, the electrostatic STM tip can be used to remove this effect, returning the QPJ to its ideal behavior.
5.3 Discussion

In this work, we propose a versatile new platform for performing electron quantum optics [10, 12, 33]. It is not hard to see how existing constructions, such as the Mach-Zehnder electron interferometer [57], can be implemented directly on the surface of an AFM TI. Fig. 5.8(a), shows a domain-wall loop channel intersecting a step channel. In this case, the interferometer works by splitting the incoming current (that flows on the step channel) in two parts that encircle the area defined by the domain-wall loop and meet at the second QPJ where they interfere. An Aharonov-Bohm phase can be introduced by threading the loop with a magnetic flux. In this way, varying the external magnetic field results in oscillations of the output conductance. The ability to control the $S$ matrix of the QPJs means they can be calibrated so that the interferometer can be used as a sensitive sensor.

In the quantum-Hall regime of 2D electron gases, the long edge-state coherence length [83, 102] and the on-demand creation of indistinguishable, single-electron WPs [11,
33, 38] have inspired ambitious proposals that consider electron interferometers as platforms for quantum information processing [13, 16, 44]. In this approach, electronic flying qubits [148] – another prominent scheme is based on photonic flying qubits [65] – are subjected to quantum operations while they are being coherently transferred, providing control over qubit separation and non-local entanglement. In contrast to photons, electrons are subject to Coulomb interactions, making them vulnerable to dephasing but, at the same time, allowing control of the entanglement strength and manipulation of the phase [17].

In the context of quantum Hall systems, entangling devices have been constructed by Coulomb-coupling two Mach-Zehnder interferometers to induce a relative phase $\phi$ between the WPs of the two coupled channels [137]. These devices can be used as electronic quantum erasers [58, 137] or even as entangling quantum gates, i.e., controlled phase gates [16]. In fact, since we have shown that QPJs can implement any single-qubit gate, a gate such as that in Fig. 5.8(b), can be adopted to perform the two-qubit entanglement for a universal set of quantum gates. We also remark that chiral Majorana fermions, first seen in magnetic TI-superconductor structures [49] and more recently in topological superconductors [60, 88, 135], are the superconducting analog of the chiral fermions discussed here. It has been proposed that topological quantum computing can be achieved using WPs propagating on chiral Majorana channels [73]. An interesting question is whether analogs of robust QPJs can be constructed in the superconducting case.

Finally, we comment on issues of temperature of operation and decoherence. In one sense, our scheme is robust to higher temperatures than quantum Hall systems; we only require the operating temperature to be small compared to the Neel temperature and the band gap in which the chiral mode is propagating. Thus, operation at tens of Kelvins is plausible for some future materials realization. However, we are keenly aware that any application that is sensitive to decoherence will require operation at lower temperatures, perhaps comparable to those needed for the quantum Hall platform, to avoid dephasing due to electron-electron, electron-phonon and electron-magnon interactions.
Nonetheless, the absence of backscattering strongly suppresses these dephasing processes relative to a normal quasi-one-dimensional wire [19] and we expect this will lead to a long edge state coherence length. In the quantum Hall context for example, edge state decoherence has been studied using Mach-Zehnder interferometry, demonstrating coherence lengths up to 100µm [54, 102]. Since each quantum operation is performed within $L \simeq 1\mu m$, it should be possible to perform many qubit operations before decoherence sets in [16]. An implementation of the Mach-Zehnder interferometer on the surface of an AFM TI would open the path to the study of coherent transport in this exciting new family of materials.

5.4 Methods

Model Hamiltonian. — We consider an adaptation of a simple four-band tight-binding model proposed by Bernevig et al [7, 53], to describe systems exhibiting a topological phase transition mediated by a single band inversion at $\Gamma$. The simplicity of the model makes detailed calculations practical even for large systems. The model is written in terms of two spinful orbitals per lattice site and takes the form

$$H_0 = m \sum_\ell c_\ell^\dagger \tau^x c_\ell + \frac{t}{2} \sum_{\ell \ell'} c_\ell^\dagger \tau^x c_{\ell'} + \frac{-i\lambda}{2} \sum_{\ell \ell'} c_\ell^\dagger \tau^x \hat{n}_{\ell \ell'} \cdot \sigma c_{\ell'} + m_Z \sum_\ell (-)^{\ell_x} c_\ell^\dagger \sigma^z c_\ell. \quad (5.11)$$

Here $\ell$ labels a lattice site $\mathbf{R}_\ell = (\ell_x, \ell_y, \ell_z)$ on the unit cubic lattice, $\sum_{\ell \ell'}$ indicates a sum over nearest neighbor sites, and $\hat{n}_{\ell \ell'}$ is the nearest neighbor unit vector. We have adopted an implied sum notation for the orbital and spin degrees of freedom, e.g.,

$$c_\ell^\dagger \tau^\mu \sigma^\nu c_{\ell'} = \sum_{ijs} c_{\ell is}^\dagger \tau_i^\mu \sigma_j^\nu c_{\ell' js},$$

where $\tau$ and $\sigma$ are Pauli matrices for orbital and spin degrees of freedom respectively, and $c_{\ell is}$ creates an electron on site $\ell$ in orbital $i$ with spin $s$.

We first consider the case where $m_Z = 0$, in which case the Hamiltonian $H_0$ reduces to the time-reversal ($T$) symmetric model proposed by Bernevig et al. [7, 53] $T$ symmetry is an axion-odd symmetry, meaning the axion coupling is quantized to $\theta_{\ell is} = 0$ or $\pi$ in its presence. At half-filling and for $(m, t, \lambda) = (1.0, -0.5, -0.6)$, the ground state has
\begin{align*}
\theta_{cs} &= \pi, \text{ corresponding to a strong topological insulator (TI). The bulk-boundary correspondence then implies the presence of } T \text{-protected surface Dirac cones, as illustrated in Fig. 5.9(d), for the (001) surface.}

\text{For } m_Z = -0.3, \text{ even though } T \text{ is broken, } T \text{ followed by a half-lattice translation along } \hat{z} \text{ (} T \ast \tau_{1/2} \text{) is a good symmetry. This symmetry is axion-odd as well [79, 127] and } \theta_{cs} = \pi \text{ for the above choice of parameters, making it an AFM-TI insulator. An important difference between } T \text{ and } T \ast \tau_{1/2} \text{ is that the latter symmetry does not force the surface AHC of all surfaces to vanish. In fact, Fig. 5.9(f), shows that the AFM Zeeman term opens the (001) surface gap, with the top and bottom surfaces exhibiting a half-integer } (n + 1/2)e^2/h \text{ AHC, where } n \text{ depends on the surface termination [126].}

\text{Wave-packet construction. — We construct the initial WPs in the space of momentum } k_{\parallel} \text{ along the direction of propagation. We calculate the surface band structure for a supercell Hamiltonian } H_{dw} \text{ or } H_{st} \text{ containing a domain wall or step, whose presence results in mid-gap bands localized on the conducting channels in the otherwise gapped}
surface, as shown in Fig. 5.1(c,d) respectively. Note that technically each slab contains two domain walls and two steps. In the domain wall case the configuration as a whole is invariant under time reversal times inversion, so the bands shown are Kramers degenerate.

For clarity we focus on the construction of domain-wall channel WPs, since the construction of step channel WPs follows in a similar fashion. We start by considering an $N_x \times 1 \times 1$ supercell of the AFM-TI model and use it to create a slab that is $N_z$ unit cells thick along (001). Because of the periodic boundary conditions, we have no choice but to create two domain walls. We choose these to be centered at the $x = 1/2$ and $(N_x + 1)/2$ planes, and create them by flipping the Zeeman potential of all orbitals in the half-cell to the left of the $x = 1/2$ plane. We refer to the Hamiltonian of this slab supercell with a pair of domain walls as $\hat{H}_{dw}$. Fig. 5.10(a), shows a sketch of the slab supercell while the associated energy bands were presented in the main text Fig.1(c) for a supercell with $N_x = 20$ and $N_z = 8$. We see that two counter-propagating, doubly degenerate, linear bands appear in the insulating gap, corresponding to the four chiral channels indicated with blue arrows in Fig. 5.10(a) (note the in-plane periodic boundary conditions).

The Bloch eigenvectors of $\hat{H}_{dw}$ are $\psi_{nk_y\sigma\tau}^{dw}(r) = e^{ik_y y} u_{nk_y\sigma\tau}^{dw}(x,z)$, where $u_{nk_y\sigma\tau}^{dw}$ are the cell-periodic counterparts. Note that we neglect the exponentially small dispersion along the $k_x$-direction, and the $y$ dependence is absent from $u_{nk_y\sigma\tau}^{dw}$ because there is only a single site per unit cell. A technical difficulty arises from the fact that the chiral channels are doubly degenerate as a result of the $I* T$ symmetry of $\hat{H}_{dw}$, but we extract channel-localized states by diagonalizing the $z$ operator in the space of the two degenerate states. We denote the channel-localized states as $\tilde{u}_{nk_y\sigma\tau}^{dw}(x,z)$, where the index $\nu$ labels the four chiral channels.

The final step in constructing our initial WPs is to take quantum superpositions of channel-localized Bloch states $\psi_{nk_y\sigma\tau}^{dw}(r)$ according to a Gaussian envelope function $F(k_y) = A \exp(-k_y^2/2\kappa^2)$ centered at $k_y = 0$, where $A$ is a normalization factor and
Figure 5.10: Wave packet construction. (a), Sketch of the AFM-TI supercell slab with a domain-wall in direction $y$ and the associated energy states. (b), Spatial profile of the transverse shape of the WP $|\chi(x, z)|^2 = \sum_{\sigma\tau} |\chi_{\sigma\tau}(x, z)|^2$. (c), Logarithmic plot of the layer density $|\chi(z)|^2 = \sum_x |\chi(x, z)|^2$ which shows the exponential localization of the states at the surface. (d), The initial (001)-projected domain-wall WP $|\Psi(x, y)|^2 = \sum_{\sigma\tau,z} |\Psi_{\sigma\tau}(x, y, z)|^2$. Note that the constructed WP (inside the box) has been placed in a larger system with zero amplitude assigned outside the box. (e-h), Same as (a-d), but for a step along the $x$ direction. Comparing (b) and (f), we see that in the case of a step the surface state wavefunction redistributes itself across the two surfaces in an unequal manner.

\(\kappa\) is a measure of the extent of the WP along the channel.\(^2\) Our initial WPs for the domain-wall are then

\[
\Psi_{\sigma\tau}(r, t = 0) = \int_{-\pi}^{\pi} dk_y F(k_y) \tilde{\psi}_{\nu k_y \sigma\tau}(r).
\]

At this point we make the additional approximation $\tilde{\psi}_{\nu k_y \sigma\tau}(x, z) = \tilde{\psi}_{\nu 0 \sigma\tau}(x, z)$, which is well justified for a WP of sufficiently narrow extent in $k_y$. We note that this approximation is equivalent to the WP decomposition of Eq. (1) in the main text with the identification

\[
\chi_{\sigma\tau}(x, z) = \tilde{\nu}_{\nu 0 \sigma\tau}(x, z), \quad f(r) = \int_{-\pi}^{\pi} dk e^{ikr} F(k).
\]

\(^2\)Note that the shape of the WP does not influence our analysis, but in more realistic applications, minimal excitations such as Levitons [56, 70] might be more desirable.
The WP construction for the case of the step channel follows in a similar way. Because the steps only need to be created at the top surface of the slab, the channel-localized states are nondegenerate. We note that the periodic boundary conditions enforce a second step channel, which produces states of both chirality in the dispersion shown in Fig. 5.1(d) of the main text. The rest of the above discussion applies, but with $x \leftrightarrow y$ and $k_x \leftrightarrow k_y$ because we take the steps to propagate along $\hat{x}$.

Wave-packet dynamics. — To avoid finite-size effects, we require the system size $L$ to be much larger than the extent of the WPs along the channel. When both a domain wall and step are present, momentum is no longer a good quantum number in any direction, so we compute the time evolution entirely in real space. This is done using Chebyshev series expansion methods [136] applied to the time-evolution operator $e^{-iHt}$. We use slabs of size $160 \times 160$ in-plane and 16 cells thick, enough to minimize finite-size effects, and adopt a Chebyshev expansion order of $N_C = 2^{11}$ so that we can time evolve the state accurately over the needed time intervals.

STM tip modeling. — To model the effects of the magnetic and electrostatic STM tips we extend the QPJ Hamiltonian ($H_{QPJ}$) with two spatially dependent terms

$$\tilde{H}_{QPJ} = H_{QPJ} + V_Z \sum_{\ell \in \Omega_Z} c^\dagger_{\ell} \sigma_z c_{\ell} + V_G \sum_{\ell \in \Omega_G} c^\dagger_{\ell} c_{\ell}, \quad (5.14)$$

where the second term modifies the Zeeman interaction in a region $\Omega_Z$ and the third term shifts the energy of all orbitals and spins uniformly inside a region $\Omega_G$.

For a positive $V_Z$ in Equation (5.14), we choose the region $\Omega_Z$ such that it restricts the sum to surface orbitals that lie within a radius $r$ of the tip, and that already experience a negative Zeeman field from the bulk Hamiltonian of Equation (5.11). Thus, $V_Z = m_Z$ is just enough to remove the Zeeman field from these sites, and $V_Z = 2m_Z$ makes the surface-layer Zeeman field equal on both sides of the domain wall or step, as in Fig. 5.4(a). We can then tune between these extremes by taking $V_Z \in [0, 2m_Z]$, thus modeling cases in which the magnetic tip has only partially reversed the surface field. Similarly, for $V_Z < 0$, $\Omega_Z$ is chosen such that the second term in Equation (5.14) is restricted to surface orbitals experiencing a positive Zeeman field in the bulk Hamiltonian.
The region of influence of the electrostatic tip, $\Omega_G$ in Equation (5.14), is defined to be a rectangle centered at $(x_0, y_0)$ relative to the QPJ and one unit cell deep, as shown by the grey shading in Fig. 5.4(d). A WP propagating for a distance $\ell$ along any domain-wall or step channel lying inside the quantum well defined by $\Omega_G$ acquires an additional phase proportional to $\ell \Delta k$, where $\Delta k$ is the shift of the Fermi wavevector of the channel. In the approximation of linear dispersion, we have $\Delta k = V_G / h \nu_F$, where $V_G$ corresponds to a local gate voltage and $\nu_F$ is the Fermi velocity (equal to $\nu_{dw}$ and $\nu_{st}$ for domain-wall and step channels respectively). Thus, the off-centering of $\Omega_G$ defined by $(x_0, y_0)$ allows us to control the travel distances $\ell$ along each of the four “legs” near the junction, introducing extra phases that are given by

$$\alpha = -\Delta k_{dw} y_0, \quad \gamma = -\Delta k_{st} x_0$$

in Equation (5.7).
Chapter 6

Searching for axion insulators

6.1 Introduction

It is truly remarkable that within two years from the theoretical proposal of strong TIs [40, 80], a prediction of their material realization [152] and a clear experimental observation [143] was achieved. In contrast, axion insulators have been much harder to realize. In fact, in the early days the community was not particularly invested in finding them since surface magnetic doping of TIs could “do the job” by opening the surface Dirac cones. For example the first experimental discovery of the quantum anomalous Hall effect was achieved when thin films of strong TIs were ferromagnetically doped [20]. However, the inhomogeneity of the magnetic dopants leads to inevitable disorder [69] and as a result the quantized response is observed at much lower temperatures than the magnetic gap and Curie temperature allow; to date the state of the art is around $\sim 1K$ [8, 21, 22, 116]. Furthermore, as we have seen in Chapter 3, various novel phenomena can occur on the surface of an axion insulator by controlling locally the surface termination. These considerations appear to have provided enough motivation to start the search for axion insulators or at least for materials that are intrinsically magnetic and have a non-trivial axion index. In this way the half-integer (or integer in thin film geometries) surface AHC could be observed at elevated temperatures, making it more relevant to applications and opening the path to applications such the quantum point junction discussed Chapter 5.

The first prediction of a magnetic TI came much more recently, when multiple groups [45, 72, 87] identified the magnetic compound MnBi$_2$Te$_4$ to have a non-trivial axion index. This van der Waals material consists of stacked layers with Mn spins aligned in-plane ferromagnetically along z while all spins flip from layer to layer. Even
though, MnBi₂Te₄ is inversion symmetric the crystal has also $T \ast \tau_{1/2z}$, i.e. time-reversal composed with a half-translation along $\hat{z}$, which is also an axion-odd symmetry. In contrast with inversion, $T \ast \tau_{1/2z}$ imposes restrictions on certain surfaces such that (100),(010) surfaces are forced to have an odd number of Dirac cones while the (001) surface can be gapped. Early work [87] showed indications of a gapped (001) surface, but it was later found that an unexpected surface Dirac cone was present inside the energy gap [48]. Recently however, the quantum anomalous Hall effect was observed in MnBi₂Te₄ [30] and tunable surface Dirac gaps of up to 100meV were observed in Sb-doped MnBi₂Te₄ [75]. An issue that appears hard to overcome is that the compound is hole-doped, making the bulk and surface slightly metallic and obscuring any topological behaviour. This issue can be partially solved in thin-film geometries where electrostatic gating controls the position of the Fermi level. Recently, two more antiferromagnetic topological insulators were predicted; EuIn₂As₂ [144] and NpBi [145] were added to the list of magnetic TI candidates but no clear signature of axion physics has been observed.

All of these call for more research into finding new material candidates. In this chapter we will discuss our efforts in predicting axion insulators using DFT calculations. In a joint effort with the Institute of Quantum Matter at John Hopkins, we collaborate with experimental physicists to predict, synthesize and measure magnetic TIs candidates. We will discuss the difficulties of the task at hand and the strategies we developed to approach it. Finally we will show some exciting results that indicate that the magnetic compounds Eu₅(Ga,Tl)₂Sb₆ are promising magnetic TI candidates. We should note that this work is still ongoing, so these results should be considered as preliminary.

---

¹ A top gate described by capacitance $c_t$ and voltage $v_t$ and a bottom gate described by capacitance $c_b$ and voltage $v_b$ control the displacement field $D$ and carrier density $n$ by $D = c_t v_t - c_b v_b$, $n = c_t v_t + c_b v_b$. Since the carrier density $n$ is a monotonic function of the Fermi level, it is used to control its position.
6.2 General approach

An axion insulator requires both the presence of magnetism and strong spin-orbit coupling. The combination of these characteristics makes their prediction a hard task. In part, this is because the computational cost of DFT calculations increases significantly when both magnetism and strong spin-orbit need to be taken into account. In addition, the sensitivity of the results to the magnetic configuration and the difficulties in predicting it from DFT mean that magnetic configurations need to be determined experimentally and then used as input in DFT calculations. This then prevents high-throughput approaches that could enable the exploration of a larger pool of candidate materials. Instead a time-consuming, case-by-case approach needs to be taken.

In addition, the very defining properties of axion insulators make it hard to realize. Since a topological insulator requires a band inversion through the presence of spin-orbit coupling, the bulk gap can be at most $\sim 1\text{eV}$ making the compound a narrow gap semiconductor. Furthermore, for a magnetic TI we require gapped surfaces and a Fermi level inside the gap in order to observe the half-integer AHC. In summary, we want insulating bulk and surfaces but the energy gap cannot be too big because of the band inversion requirement.

Fortunately, nowadays material databases, such as the Inorganic Crystal Structure Database (ICSD), offer hundreds of thousands of crystals to search through. In fact, the recent advancements in determining topological invariants through elementary band representations \cite{15} have enabled high-throughput prediction of $\mathcal{T}$-symmetric topological materials using ICSD compounds as input \cite{128, 129}. This impressive achievement illustrates that topological materials are really not that uncommon. Indeed, in the Topological Quantum Chemistry Database (TQCD) more than 60,000 compounds have been classified based on their topological properties as predicted by DFT, indicating that around 53\% of all stoichiometric materials in nature are in some sense topological\textsuperscript{2} at intrinsic filling \cite{129}. We should note that an impressive high-throughput search

\textsuperscript{2}The definition used here is that band representation is non-trivial. This includes both insulator and enforced semimetals. However, the energetics of the band structure are not taken into account in the classification which is very important for our search.
for antiferromagnetic topological materials (≈ 500 compounds) has been recently conducted [145], predicting NpBi as an AFM TI.

We use these databases as a starting point to construct lists of promising axion insulator candidates. Specifically, we search through the ICSD and the TQCD to identify potential candidates for further investigation. The general search parameters we based our search on are:

- We require that potential candidates contains at least one magnetic element, e.g. Eu, Mn etc.
- In cases we know the magnetic configuration, we require that at least one axion-odd symmetry is contained in the magnetic space group.
- In cases we do not know the magnetic configuration, centrosymmetric materials are preferred since inversion is usually preserved when the configuration is simple enough.
- Chemical considerations should indicate that the system is insulating, stable and safe to synthesize; for example oxides or materials that satisfy the Zintl concept [84].
- Heavy elements with large spin-orbit coupling are preferable in order to open larger energy gaps.
- Compounds in the TQCD showing non-trivial topology (in the non-magnetic DFT calculations) but containing magnetic elements.

After a list of promising candidates is constructed, we proceed in coordination with our experimental collaborators to conduct careful DFT calculations, crystals synthesis and characterization.

6.3 Magnetic topological insulators in Zintl compounds

It is not hard to notice that many of the criteria outlined above match the properties of a different group of materials, the thermoelectrics. These are materials with the ability
to generate an electrical voltage when subjected to a temperature gradient (and vice versa). Some of the best performing thermoelectrics are doped intermetallic narrow-gap semiconductors [117] such as PbSe and Bi$_2$Te$_3$. These materials happen to also be ideal TIs and it is therefore natural to look for topological materials in the thermoelectric literature. A particular prominent class of thermoelectric materials are the so called Zintl compounds [74]. These are defined as valence precise intermetallic phases in which electropositive cations donate electrons to covalently bonded polyanions [59]. In the limit that electron transfer is complete a Zintl should be a semiconductor but in reality there is electron competition between cations and anions. In this way, Zintl compounds form complex crystal and electronic structures. Numerous such compounds have already been synthesized and characterized, featuring new structure types and unusual bonding patterns, along with a variety of interesting physical properties including superconductivity [29], magnetoresistance [18, 101] and thermoelectricity [74]. This makes Zintl compounds a promising searching ground for topological insulators.

Of particular interest in our search for axion insulators are Zintls where the role of the cation is played by a divalent Eu$^{2+}$. In this case the magnetism comes from the localized, spin-polarized $f$-orbital manifold, which implies that crystal-field and Kondo effects are negligible ($J = S = 7/2$). Furthermore, the occupied and unoccupied $f$-electrons are expected to be separate by a strong Hubbard $U \sim 5$eV so that they lie far from the Fermi level. In this way they act as a magnetic perturbation that is not strong enough to cause a band inversion but will in principle gap any surface Dirac cone. This enables us to first analyze the band structure and topological properties in a non-magnetic calculation, i.e., by placing the Eu$^{2+}$ $f$-electrons in the core, and if there are signs of insulating behaviour and non-trivial topology, then perform the more time-consuming magnetic calculation.

Excitingly, this approach was recently validated by Xu et al. [144] with the prediction of a magnetic TI candidate in the Zintl compound EuIn$_2$As$_2$. The space group contains inversion, so that the non-trivial axion $Z_2$ index can be calculated using the Fu-Kane criterion. The lowest energy magnetic configuration was predicted to be A-type AFM, as in MnBi$_2$Te$_4$, with the important difference that the moments are pointing in-plane.
In this case the magnetic symmetries force the (100), (010) and (001) surfaces to host Dirac cones, making the compound a less viable candidate to observe the half-quantized physics we are looking for. However, recent work \cite{100} used neutron diffraction and DFT calculations to show that in fact EuIn$_2$As$_2$ exhibits a low-symmetry helical antiferromagnetic order. This magnetic configuration breaks inversion, but $C_2 \ast T$ along the (110) and (110) directions are good symmetries, making it a magnetic TI.

As part of our search, we have performed DFT calculations and symmetry analysis of the Eu based Zintl compounds shown in Table 6.1. Except for the the already known candidate EuIn$_2$As$_2$, none of the existing Zintls we have looked at was a magnetic TI. However, in the following we predict that the not-yet-synthesized Zintl compounds Eu$_5$(Ga,Tl)$_2$Sb$_6$ should be magnetic TIs.

Table 6.1: Characterization of Zintl compounds containing Eu$^{2+}$ based on our DFT calculations. Left column “122” and right column “526” structures.

<table>
<thead>
<tr>
<th>EuMn$_2$P$_2$ \cite{90}</th>
<th>Trivial insulator</th>
<th>Eu$_5$Al$_2$Sb$_6$ \cite{94}</th>
<th>Trivial insulator</th>
</tr>
</thead>
<tbody>
<tr>
<td>EuMn$_2$As$_2$ \cite{103}</td>
<td>Topological semimetal</td>
<td>Eu$_5$In$_2$As$_6$ \cite{25}</td>
<td>Trivial insulator</td>
</tr>
<tr>
<td>EuIn$_2$As$_2$ \cite{144}</td>
<td>Topological insulator</td>
<td>Eu$_5$In$_2$Sb$_6$ \cite{89}</td>
<td>Trivial insulator</td>
</tr>
<tr>
<td>EuCd$_2$Sb$_2$ \cite{4}</td>
<td>Trivial insulator</td>
<td>Eu$_5$Sn$_2$Sb$_6$ \cite{132}</td>
<td>Trivial insulator</td>
</tr>
<tr>
<td>EuCd$_2$As$_2$ \cite{4}</td>
<td>Topological semimetal</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

6.3.1 Prediction of magnetic topological insulator in Eu$_5$(Ga,Tl)$_2$Sb$_6$

The Zintl compound Eu$_5$In$_2$Sb$_6$ has been known for sometime now \cite{89}. It is a centrosymmetric narrow-gap antiferromagnetic semiconductor which crystallizes in an orthorhombic structure (space group 55). Recently Rosa et al. \cite{101} investigated its properties with an emphasis on the observed colossal magnetoresistance. In addition, they performed DFT calculations within the generalized gradient approximation (GGA) with and without using the modified Becke-Johnson potential. In the former case and in the paramagnetic phase, i.e., Eu $f$-electrons set in the core in the core, they predicted an indirect gap and a non-trivial $Z_2$ index. The situation did not change after they considered various putative antiferromagnetic configurations within GGA + U,
Figure 6.1: (a) Using the correct crystal structure, obtained from Park et al. [89], DFT predicts the Zintl compound Eu$_5$In$_2$Sb$_6$ to be a trivial insulator with a direct band gap. (b) Rosa et al. [101] predicted a non-trivial $Z_2$ index due to the displacement of the In atoms along the y direction.

suggesting that the antiferromagnetic phase of Eu$_5$In$_2$Sb$_6$ may host an axion insulating state.

Meanwhile, as part of our search for axion insulator candidates in the Zintl, our DFT calculations, within the GGA approximation, classified Eu$_5$In$_2$Sb$_6$ as a trivial insulator with a narrow band gap around 50meV. After contacting the researchers involved in the DFT calculations of Rosa et al. [101], we realized that the discrepancy between our results, resulted from a mistake they had made in one of the atomic coordinates of the crystal. Specifically, the $y$ component of the Wyckoff position of the In atom, which is not pinned by symmetry, was incorrect due to a transcription error.\footnote{$y_n = 0.2418$ in their case, and $y_n = 0.2149$ in our case.} As Fig. 6.1(b) shows, the displacement of the In atom modifies the In – Sb bond length...
Figure 6.2: Prediction of non-trivial $Z_2$ index in the paramagnetic phase of (a) Eu₅Ga₂Sb₆ and (b) Eu₅Tl₂Sb₆.

which in turn results in band inversion at the $\Gamma$ point.

Even though the exact mechanism driving the band inversion is not yet clear, this result clearly indicates that the In-Sb bond length can drive a band inversion. This motivated us to substitute In with elements in the same column of the periodic table i.e., Ga and Tl.\textsuperscript{4} Since Eu₅Ga₂Sb₆ and Eu₅Ga₂Sb₆ After relaxing the internal parameters and bond lengths we find that both compounds remain in the same space group. In addition, for both Ga and Tl the environment remains approximately tetrahedral. However, there is a 12% decrease and 16% increase in the bond length in the Ga-Sb and Tl-Sb bonds respectively. Fig. 6.2 shows that both substitutions result in band inversions and a non-trivial $Z_2$ index in the paramagnetic phase, i.e., when Eu $f$-electrons are set in the core.

Until this point we have neglected the magnetic moments of Eu$^{2+}$ in order to identify the topology. That was based on the valid assumption that their effect is to perturb

\textsuperscript{4}Note that Eu₅Al₂Sb₆ [94] has been synthesized and crystallizes in space group 62 instead of 55.
the states close to Fermi level so that the $Z_2$ index will remain unaffected. However, knowledge of the magnetic symmetry is vital for determining the surface properties of magnetic TIs. In the following section we theoretically investigate the complex magnetic configuration of $\text{Eu}_5(\text{Ga,In,Tl})_2\text{Sb}_6$. Since all three compounds follow the same trends we will focus $\text{Eu}_5\text{In}_2\text{Sb}_6$, but the same conclusions apply to all of them.

### 6.3.2 Determination of the magnetic configuration

The compound has a layered structure along the $c$ direction. The atoms are localized on two planes, “integer” planes containing Eu and Sb and “half-integer” planes containing In and Sb. For the purpose of the magnetic configuration, we can therefore consider only the integer planes containing Eu. Fig. 6.3 shows an integer plane with three symmetry-independent Eu sites shown with different colors. One site corresponds to Wyckoff position 2a, which we refer to as Eu2, and the other two sites have Wyckoff position 4g, which we refer to as Eu1 and Eu3.

The non-magnetic space group 55 consists of 8 elements, three of which are mirror symmetries, i.e. $\{m_a|1/2,1/2,0\},\{m_b|1/2,1/2,0\}$ and $\{m_c|0,0,0\}$. With $\mathcal{T}$ broken the non-magnetic symmetries of space group 55 need to be composed with $\mathcal{T}$ in order to remain good symmetries. This gives rise to 8 irreducible representations that can be understood as the $2^3$ combinations from the composition of $\mathcal{T}$ with the three mirrors. Table 6.2 gives a summary of the 8 irreps. Notice that we have divided the irreps into two groups depending on whether the spins are pointing in the $ab$ plane or along the $c$ direction. Initial neutron scattering data from $\text{Eu}_5\text{In}_2\text{Sb}_6$ indicates that the spins are aligned in plane so that in the following only the four irreps will be considered. Finally, we consider the possibility of doubling of the unit cell along the $c$-direction with the spins reversed from layer to layer, which we will refer to these irreps as $Z(\text{irrep} \#)$. When the magnetic unit cell remains the same as the non-magnetic, we refer to as $\Gamma(\text{irrep} \#)$.

Importantly, $\text{Eu}_5\text{In}_2\text{Sb}_6$ undergoes two magnetic transitions as can be seen in heat capacity measurementsRosa et al. [101]. In addition, neutron scattering indicates that
Table 6.2: Spin irreducible representations for Eu526-class materials. Irrep #: irreducible representation labels. Mirror: presence of mirror $m_i$ or time-reversed mirror $m'_i$ along $i = a, b, c$ directions. Spin: Direction along which spins are pointing at. Mag: magnetization direction, if any. 2a: Absence/presence of the corresponding irrep for Eu with Wyckoff position 2a. 4g: Same for Wyckoff position 4g.

<table>
<thead>
<tr>
<th>Irrep #</th>
<th>Mirror</th>
<th>Spin</th>
<th>Mag</th>
<th>2a</th>
<th>4g</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>$m_aim_bm'_c$</td>
<td>$ab$</td>
<td>$a$</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>5</td>
<td>$m'_aim_bm'_c$</td>
<td>$ab$</td>
<td>$b$</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>2</td>
<td>$m'_aim'_bm'_c$</td>
<td>$ab$</td>
<td>-</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>8</td>
<td>$m_aim_bm'_c$</td>
<td>$ab$</td>
<td>-</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>1</td>
<td>$m_aim_cm_c$</td>
<td>$c$</td>
<td>-</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>7</td>
<td>$m'_aim'_bm_c$</td>
<td>$c$</td>
<td>$c$</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>4</td>
<td>$m'_aim_cm_c$</td>
<td>$c$</td>
<td>-</td>
<td>✗</td>
<td>✓</td>
</tr>
<tr>
<td>6</td>
<td>$m_aim'_bm_c$</td>
<td>$c$</td>
<td>-</td>
<td>✗</td>
<td>✓</td>
</tr>
</tbody>
</table>

during the first magnetic transition between 7-14K the unit cell is preserved, i.e. Γ-ordering, while below 7K the unit cell doubles along the $c$-direction. It is believed that Eu2 order first at 14K followed by Eu1 and Eu3 at 7K. Based on these experimental evidence, we begin with the 7-14K region, which we model by treading Eu1 and Eu3 as non-magnetic, and include the $f$-electrons in the valence only for Eu2.

The DFT-predicted magnetic structure between 14K and 7K is shown in Fig. 6.3(a). Note that our calculations, not shown here, indicate that Z-ordering is not favored so the unit cell should remain unchanged in agreement with experiment. Since we only consider spin irreps that order in-plane, all possible configuration are described by two angles $\theta_1, \theta_2$. For Eu2 there are only two available irreps, i.e., irrep 3 and irrep 5, and can be parametrized by single variable $\theta$ as

$$\text{irrep 3} : \theta_1 = 180^\circ + \theta, \theta_2 = -\theta$$

$$\text{irrep 5} : \theta_1 = 90^\circ - \theta, \theta_2 = 90^\circ + \theta.$$  \hspace{1cm} (6.1)

Notice that when $\theta = 0$ both irreps correspond to AFM configurations and the DFT
Figure 6.3: (a) Integer-value layer containing the three symmetry independent Eu sites. The figure illustrates the predicted magnetic configuration between 7K and 14 K which is AFM along the a-direction with a small canting of $4.2^\circ$ towards the b-direction. (b) DFT calculation of the energy associated with magnetic configurations of $\Gamma(3)$ and $\Gamma(5)$ as defined by Eq. (6.1). The system favors AFM configurations and alignment along the a-direction. (c) Contour plot of Eq. (6.2) with the minimum configuration indicated with a red star.

calculations in Fig. 6.3(b) show that they are energetically preferable. In contrast, when $\theta = 90^\circ$ both irreps are in FM configurations, which is the least favorable. These considerations indicate that the biggest energy scale is the in-plane exchange interaction $J_\parallel \approx 1\text{meV}$, while anisotropy is an order of magnitude smaller, $A \approx 0.1\text{meV}$. To quantify these trends we consider the energy of the system as a function of the angles $\theta_1$ and $\theta_2$, modeled as

$$E(\theta_1, \theta_2) = A \cos 2(\theta_1 + \phi) + A \cos 2(\theta_2 + \phi) + J \cos \theta_2 - \theta_1.$$  \hspace{1cm} (6.2)

Here the in-plane exchange interaction $J_\parallel$, the anisotropy $A$, and easy axis $\phi$ are fitting parameters. By fitting this model to our DFT calculations, we find $J_\parallel = 1.23\text{meV}$, $A = 0.11\text{meV}$ and $\phi = 31^\circ$. Fig. 6.3(c) shows a contour plot of Eq. (6.2) with the minimum at $(\theta_1, \theta_2) = (+85.8, -85.8)$ which corresponds to irrep 5 and is shown in Fig. 6.3(a). In summary, we predict an almost AFM configuration, but the spins are canted by $4.2^\circ$ from the a-direction due to the anisotropy, resulting in a small non-zero magnetization along the b-direction.

The magnetic configuration below 7K, when all Eu atoms are expected to order, is harder to predict, since the available “phase space” is much larger. To proceed we
Figure 6.4: (a) Predicted magnetic configuration below 7K is A-type AFM with in-plane moments. Small canting along the $b$-direction is also expected. (b) Energies of different magnetic configuration when only Eu1 and Eu3 $f$-electrons are included in the calculation. (c) The band structure and topological index of Eu$_5$Ga$_2$Sb$_6$ in the A-type AFM configuration are unaffected compared to the paramagnetic calculation.

will assume that Eu1 and Eu3 will occupy the same spin irrep. It is instructive to set Eu2 $f$-electrons in the core and consider the four irreps where the spins lie in the $ab$ plane. Fig. 6.4(b) shows the energies of the four in-plane spin irreps, both for $\Gamma$ and $Z$ ordering. The affect of anisotropy is also considered by globally rotating all spins from the $a$ axis to the $b$ axis. The slightly shifted points with the same color correspond to spins pointing along $b$. The energy scale of the anisotropy is much smaller, and it is expected, just like the Eu2 case, that the otherwise collinear spins will be slightly canted, resulting in a small magnetization along the $b$-direction. Therefore in the following we neglect their effects and consider collinear configurations.

The first thing to notice is that the energy scale for the exchange interaction in Fig. 6.4(b) is much larger than that of Eu2 in Fig. 6.3(b). This makes Eu1 and Eu3 the determining factors for the lowest energy configuration, and indeed when Eu2 $f$-electron are also added, the picture does not change. Now if we notice that irrep 2 and 8 are locally AFM while irrep 3 and 5 are locally FM; then Fig. 6.4(b) tells us that the system wants to be either $\Gamma$-ordered, i.e., FM along $c$ and AFM in-plane or $Z$-ordered, i.e., AFM along $c$ and FM in-plane.

Even though, $\Gamma'(2)$ is the lowest energy in Fig. 6.4(b), when Eu2 are include, $Z(3)$ is the lowest in energy for Eu1, Eu2 and Eu3. This can be understood because Eu2 cannot be in irrep 2, see Table 6.2. Therefore the system lowers its energy by setting
Eu1, Eu2 and Eu3 in \(Z(3)\), i.e., A-type AFM with the moments along the \(a\) axis, which results in a highly symmetric magnetic space group.

Finally, the band structure calculations and symmetry analysis of \(\text{Eu}_5\text{Ga}_2\text{Sb}_6\) and \(\text{Eu}_5\text{Tl}_2\text{Sb}_6\) in the A-type AFM configuration show that both remain magnetic TIs as with our expectation.

### 6.4 Conclusions

In this chapter, we described our efforts to find magnetic TIs using DFT calculations. To do that we have outlined certain criteria that any potential candidate is required to have, and some others that would be favorable if present. We then searched through material databases to find candidates and perform more detailed and systematic calculations. One family of materials that is especially promising is that of Zintl compounds where the role of the cation is played by a divalent Eu atom. Our initial calculations of promising Zintls was not fruitful, but due to a discrepancy between our calculations and those of Rosa et al. [101] regarding \(\text{Eu}_5\text{In}_2\text{Sb}_6\), we found a chemical mechanism that could drive a band inversion. This then led us to predict \(\text{Eu}_5\text{Ga}_2\text{Sb}_6\) and \(\text{Eu}_5\text{Tl}_2\text{Sb}_6\) as magnetic TIs. Finally, we performed DFT calculations and analysis to uncover the complex magnetic structure of these compounds.

At the time that this thesis is being written, this work is still ongoing, so that many questions are still open. For example, the chemical mechanism driving the band inversion is not yet clear. Identifying such a mechanism could help us engineer band inversions in Zintl compounds and accelerate the search. More importantly, \(\text{Eu}_5\text{Ga}_2\text{Sb}_6\) and \(\text{Eu}_5\text{Tl}_2\text{Sb}_6\) have not yet been synthesized, and it is not clear whether this is an easy task. For example, the element Tl is not an easy chemical to work with, as it can be toxic. Finally, regarding the magnetic configuration of the 526 compounds, there is ongoing experimental work, i.e. neutron scattering on \(\text{Eu}_5\text{In}_2\text{Sb}_6\) which indicates possible non-collinear order. More feedback from experiment is expected to clarify and explain the complex magnetic structure of the 526 compounds.
Chapter 7

Summary and outlook

In this work we have explored the physics of the axion coupling with an emphasis on magnetic TIs. These are crystals with symmetries other than simple time-reversal, that quantize the axion coupling and have $\theta_{cs} = \pi$.

We made significant contributions by elucidating their physical properties and mathematical manifestation. In Chapter 3, we formally answered the question of how the integer part of the surface anomalous Hall conductivity is determined. This then enabled the discovery of novel phenomena at the surfaces of axion insulators. In Chapter 4, we considered how the axion coupling manifests in the hybrid Wannier representation. After describing the general features, we classified all axion-odd symmetries and explained how each class manifests.

Furthermore, we proposed a novel device that harnesses the unique properties of axion insulators and predicted material realizations. In Chapter 5, we took a closer look at the quantum point junction proposed in Chapter 3, and showed how it realizes a robust and controllable quantum gate with applications in electron quantum optics. Finally, in Chapter 6, we discussed our search for an axion insulator using DFT, and presented our preliminary results that suggest the Zintl compounds Eu$_5$(Ga,Tl)$_2$Sb$_6$ as potential candidates.

Looking at the big picture, in recent years the community has realized the potential of magnetic topological insulators and a lot of developments have occurred. Perhaps the most important of these was the discovery of an antiferromagnetic topological insulator in MnBi$_2$Te$_4$ [87]. Given the progress, such as the discovery of the anomalous Hall effect observed in MnBi$_2$Te$_4$ [30] and tunable surface Dirac gaps of up to 100meV observed in Sb-doped MnBi$_2$Te$_4$ [75], the material quality should only be expected to improve.
This will open the path to using the material in the plethora of applications proposed. As was discussed in detail in Chapter 5, this has many technological benefits, ranging from chiral channels in the absence of an external magnetic field, to higher temperature manifestations of the phenomena. However, the hallmark of axion electrodynamics, i.e., the topological magnetoelectric effect, is still far from being realized as it cannot be realized in MnBi$_2$Te$_4$ [30], where only the (001) surfaces are gapped. Indeed, for the topological magnetoelectric effect to occur one needs “many” gapped surfaces in order for the Hall current to circulate around the bulk of the material. This then creates a magnetization parallel to the external electric field with a constant of proportionality $\alpha = e^2/2\hbar$. In order to achieve this one needs an AFM TI with a complex antiferromagnetic configuration and control over the surface termination so that all surfaces have the same sign of the AHC, i.e., inwards or outwards. Beyond the topological magnetoelectric effect, even more exotic phenomena are possible if one can realize the dynamical axion field $\theta_{cs}(r,t)$. This could be achieved spontaneously when the Weyl cones of a Weyl semimetal are gapped out due to the presence of a charge-density wave [133], or perhaps by perturbing the axion field externally, e.g., through a quantum quench or a periodic driving force, to produce ripples in the otherwise constant axion field of the axion-odd insulator.
Bibliography


[113] J. C. Y. Teo, L. Fu, and C. L. Kane. Surface states and topological invariants in


