Title
Electric and Magnetic Response Properties of Topological Insulators in Two and Three Dimensions

Author
Essin, Andrew

Publication Date
2010

Peer reviewed|Thesis/dissertation

Electric and Magnetic Response Properties of Topological Insulators in Two and Three Dimensions

by

Andrew Michael Essin

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Physics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Joel E. Moore, Chair

Professor Ashvin Vishwanath

Professor Yuri Suzuki

Spring 2010

Electric and Magnetic Response Properties of Topological Insulators in Two and Three Dimensions

Andrew Michael Essin

Abstract[1]

Electric and Magnetic Response Properties of Topological Insulators
in Two and Three Dimensions

by

Andrew Michael Essin

Doctor of Philosophy in Physics

University of California, Berkeley

Professor Joel E. Moore, Chair

This dissertation introduces some basic characteristics of a class of materials known as topological insulators. These materials were introduced theoretically as crystalline band insulators, in which electrons do not interact with each other and the atomic cores form a perfectly ordered, fixed background potential for the electrons. It is shown that, in the two-dimensional case, this definition can be relaxed to the case of disordered, noninteracting insulators. It is further shown numerically that the introduction of disorder to these two-dimensional insulators eliminates the direct transition between the topological and ordinary insulating phases, consistent with the presence of an intervening metallic phase.

In the three-dimensional case, one formulation of the distinction between the topological and ordinary phases involves a quantized response function (the magnetoelectric polarizability). It is shown here that this characterization holds on quite general grounds, and therefore allows an extension of the topological class to disordered and interacting systems. Finally, a relatively rigorous derivation of the (fixed-ion, linear, orbital) magnetoelectric response of crystalline band insulators is presented.

[1] Yes, but some parts are reasonably concrete.
— Avron *et al.*, Commun. Math. Phys. **123**, 595 (1989)

Contents

List of Figures

List of Tables

Acknowledgments

This thesis is primarily a record of some of the research I have worked on during my time as a graduate student at UC Berkeley, and much of the credit for the results presented, and for whatever success I have had, must rest with others.

Most significantly, Professor Joel Moore has suggested the problems I have studied and given me the background I needed as well as the methods, and sometimes even the computer code. Even more importantly, he has helped me to work through my shortcomings as a researcher when appropriate, and also has tried to show me when to abandon an unproductive line of work.

Besides Joel, my collaborators David Vanderbilt and Ari Turner deserve much of the credit for the results presented in this dissertation. Indeed, Ari is largely responsible for pushing me to actually work on the fully quantum treatment of the orbital magnetoelectric response, and he was the one to show that a derivation from density-matrix perturbation theory is possible. In addition to his concrete contributions to this research, David's alternative perspective on the interesting and important features of condensed matter systems was uniformly enlightening.

I also want to thank Andrei Malashevich and especially Ivo Souza for many discussions about magnetoelectric effects and the derivation of the response function. Although we eventually wrote two separate papers on the topic rather than one gigantic paper, they were still instrumental to the progress of the research. Andrei, one of the world's two experts in the numerical computation of orbital magnetoelectric responses, was gracious enough to compute the responses in a model I cooked up that was too challenging for my numerical abilities.

Ivo deserves further thanks, along with Roger Mong and Marcel Franz, for bringing to my attention numerical errors in the published version of the results presented here in Chapter 4.

Thanks to Hal Haggard for fielding my random questions about various topics in physics and mathematics, and for giving a careful reading to a part of this thesis. Thanks are also due to Alex Selem, with whom I had a couple of productive discussions that ultimately led to the model Hamiltonian presented in Chapter 4.

Finally, I need to thank Joel and Professors Joe Orenstein, Yuri Suzuki, and Ashvin Vishwanath for their tolerance of my weak performance during my qualifying exam, and their belief that I had potential in spite of that. Thanks also to Professor Suzuki for helpful comments on the first chapter of this manuscript.

Some of the work presented here has been published by the American Physical Society. The contents of Chapter 3 appeared in *Phys. Rev. B* **76**, 165307 (2007). The contents of Chapter 4 appeared in *Phys. Rev. Lett.* **102**, 146805 (2009). The contents of Chapter 5 have just appeared in *Phys. Rev. B* **81**, 205104 (2010).

Thanks to the Army Research Office for funding part of my graduate education through a National Defense Science and Engineering Graduate Fellowship, and thanks also to the Western Institute for Nanoelectronics for my last two years of research support. The computations in Chapter 3 were done on the SUR cluster donated by IBM.

Chapter 1

Background and overview

One of the major driving motivations in condensed matter physics is the search for new phases of matter, that is, new materials with novel properties. Sometimes these new properties lead directly to important technologies. Sometimes they "merely" illuminate previously obscure possibilities in the structure of the laws that govern the natural world. Sometimes they even manage to do both.

The work I will present in the following chapters grows out of the recent theoretical and, increasingly, experimental development of a new class of materials termed "topological insulators". In some sense the name is misleading, since there have been systems deserving of the name for nearly thirty years, and in principle the new discoveries form only a subset of all topological insulators. In particular, the new materials are nonmagnetic topological band insulators.

In the rest of this chapter, I want to introduce the key features of topological insulators, or at least those that are not covered in greater depth later on. Before doing so, it seems worthwhile to start at a relatively basic level and review the notion of a band insulator (and the contrasting notion of a metal) and that of a topological property.

1.1 Band theory of solids

In middle school, I was taught that materials are divided into two classes: conductors (metals) that carry electricity, and insulators that don't. The work I will be presenting deals with new kinds of insulators (and a little about metals), so I want to review a simple picture of how insulating or conducting behavior comes about, to serve as a reference point that highlights why the new materials are interesting.

The electronic spectrum of an atom is divided into two pieces, the bound states and the "scattering" (unbound) states. The bound state spectrum of an atom consists of levels labeled by n, the "principal quantum number", which are separated by energy gaps. This discrete energy spectrum is a consequence of, or is at least related to, the localized nature of the electron wave functions. An atom also has a "scattering" spectrum, consisting of those

states that are not bound to the atom. Qualitatively, these are extended free-particle states that get modified slightly by the presence of the nucleus.

Is the spectrum of a solid more like the bound-state state spectrum of an atom or the continuum? Are the states localized or extended? On the one hand, a solid is frequently conceived as being built of atoms, which suggests the former; on the other hand, thinking about nuclei and electrons rather than atoms, the (relatively) tightly packed nuclei might just provide an environment in which the electrons can run around.

The simplest picture starts from atoms. Consider a bunch of isolated atoms, all of the same species for true simplicity. Because electrons are fermions and therefore obey the Pauli exclusion principle (each state can only accommodate one electron at a time), they fill up the lowest few states of the discrete atomic spectrum (at zero temperature, which is the situation for all the work I will present here). In this situation, the electrons are localized, that is, not free to run around in the system. In some sense this is like the vacuum itself, which also has no particles. If we imagine bringing the atoms closer together, the small overlap in the wave functions of two atoms will allow the electrons to tunnel back and forth. This also produces an energy splitting, but the splitting is much smaller than the atomic energy spacing. In this picture, each atomic level broadens into an energy "band", see Figure 1.1.

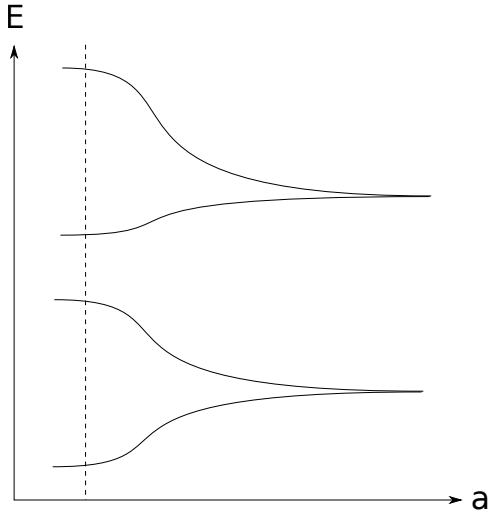

Figure 1.1: Energy bands connect continuously to the limit of separated atoms, where the atomic spacing a tends to ∞. At finite separation the hybrid orbitals develop a spread in energy ("bandwidth"). The dashed line indicates the physical spacing in a material.

Now we can try to answer the question, Does the material conduct or not?; Are the electrons free or bound? Each band n (generalizing the principal quantum number) is essentially a continuum of states, in which a traveling wave packet can exist. However, if a band is filled, there are no states available from which to assemble such a packet. An unfilled band should be conducting, while a filled band will be insulating. In this way, the band theory of solids gives a simple way to understand metals and insulators; metals have partially filled

bands, while insulators have filled bands.

Of course, this simple picture cannot account for all metallic and insulating materials. Most importantly, it assumes that the electrons do not interact with each other. Since electrons repel, a situation can arise in which electrons localize (stop conducting) so as to minimize the repulsive energy, even though the band theory would predict metallic behavior.

The topological insulators that are the primary subject of the work I will present are band insulators, but with a difference. In particular, assuming that no magnetism develops (technically, that time reversal symmetry is unbroken), the band gap will have to close upon trying to take the (isolated) atomic limit of a topological insulator. When the band gap closes, the system becomes a metal, and therefore fundamentally different from an insulator (at least in its transport properties). In this sense, the topological insulator is a distinct phase of matter from the "ordinary" or "trivial" insulator. It should be noted, however, that time-reversal breaking allows the possibility of an atomic (or at least molecular) limit. In this sense, the phase transition is more real than the phases themselves, rather like the transition between liquid water and water vapor.

1.2 Insulators, metals, and quasiparticles

In the case of metals, I said above that partially filled bands can support traveling wave packets. In such a situation, the dispersion relation between energy and momentum is crucial for understanding the wave packet motion. Indeed, it is typically more intuitive to think of particles of definite momentum and energy when possible, even though the associated wave functions will not be normalizable in an infinite system. Now, the atomic cores in a solid exert electrostatic forces on the electrons and break conservation of momentum, but a metal will still have sharply defined excitations with good energy and momentum quantum numbers; this is Landau's notion of the "quasiparticle", essentially an electron but with a different dispersion relation (and also a finite lifetime).

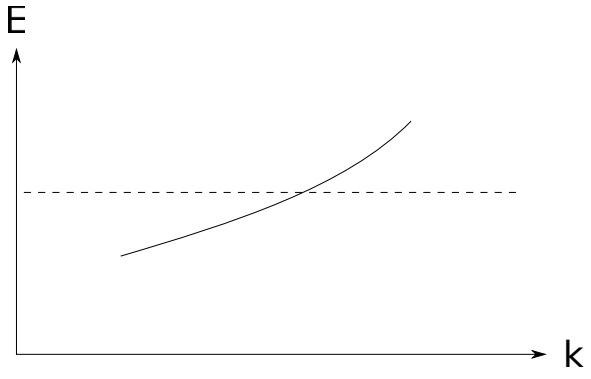

Figure 1.2: A generic quasiparticle dispersion in a metal. The dashed line indicates the chemical potential, below which all states are filled at zero temperature.

An insulator does not have quasiparticles, almost by definition: quasiparticles give us a microscopic picture for the mechanism of conduction. I should note here that I am using the term conduction in a particular sense, namely that of dissipative transport. This is tied up with the particle-like picture: charged particles respond to a small electric field by running down the potential, thereby exchanging energy with the external field source (dissipating energy, from the point of view of the material) and carrying a current. Insulators, lacking quasiparticles, cannot do this. However, they could still support dissipationless transport, which would necessarily run perpendicular to the applied field. As I will discuss soon, this is precisely what happens in the quantum Hall effect.

Finally, it is important to recognize that a material that is insulating in its bulk could have metallic surfaces (really boundaries, since I will also be discussing two-dimensional systems whose boundaries are edges rather than surfaces). Indeed, this is precisely what happens in the topological insulators, although it is also possible to conceive of topological insulators for which this is not the case.

1.3 Topological properties

The systems of primary interest in this work are topological insulators. The term "topological" is used in a number of ways in condensed matter physics, and while many of them are related they are not precisely identical, so it is worth giving a brief survey of meanings.

Topology is a mathematical notion of continuity. For a physicist, the relevant distinction is between geometry, which has to do with shape and size, and topology, which is more basic. In a popular and useful formulation, two spaces that can be deformed continuously into one another are said to be topologically equivalent, irrespective of the details of shape and size. Qualitatively, topological properties are then those that are insensitive to continuous changes.

One definition of a topological property, then, is a property that does not respond to continuous changes. Typically this is used in a quantitative sense (the winding of the superfluid phase around a minimal vortex is always 2π) rather than a qualitative one (away from a melting transition, changing the temperature of a solid does not change its solidity).

A related, perhaps more precise meaning, has to do with topological invariants. These objects characterize the topology of mathematical spaces, and sometimes arise more or less naturally in physics applications. For example, the superfluid winding mentioned above determines an integer element of the "fundamental group" of the circle parameterized by the phase; the minimal vortex corresponds to the unit element of the group of integers $\mathbb{Z}$. This meaning tends to imply the other. In the winding example, an integer is "topological" in the sense that it cannot respond to continuous changes, or any changes really, continuously.

Another related notion is that of topological order. There is no single meaning of topological order, but it typically means that the important properties of the system in question cannot be captured by any short-ranged probe or measurement. Instead, they can only be seen in quantities sensitive to the system as a whole, such as the dependence of the ground state degeneracy on boundary conditions (open or periodic, say). The degeneracy is again

an integer, which makes this topological in the above sense. This example also illustrates the sort of drastic change, like cutting open a boundary, needed to change a topological property.

The topological insulators that are the subject of this dissertation fit most naturally into the definition in terms of topological invariants, but all the characterizations above work to greater or lesser degree.

I will review some of the known properties of topological insulators in this introduction and in the subsequent chapters. For more complete coverage, I refer the reader to a few good, general reviews of the subject that have appeared recently. Moore [58] and Qi and Zhang [68] have written reasonably qualitative introductions for a general audience, while Hasan and Kane [31] have written a detailed review of the theoretical and experimental state of the art.

1.4 The quantum Hall effect

The prototypical topological systems are those that exhibit the quantum Hall effect. The effect was originally seen in the two dimensional gas of electrons that forms at the boundary between GaAs and AlGaAs, and has now also been seen in graphene, the truly two dimensional crystalline form of carbon. In sufficiently clean samples, at low enough temperatures,[1] and in a strong enough magnetic field perpendicular to the plane of the electrons, it is found that the (longitudinal) conductance σ^{xx} vanishes, while the (transverse) Hall conductance $\sigma_H = \sigma^{xy}$ takes values that are integer multiples, or simple fractions, of the natural unit of conductivity e^2/h, where e is the electron charge and h is Planck's constant.[2] Since the Hall conductance takes its values in a discrete set under these conditions, it qualifies as a topological property by the above definitions.

The crucial prerequisites for quantum Hall physics are the dimensionality (two dimensions, specifically) and the breaking of time reversal symmetry $\mathcal{T}$ supplied by the strong magnetic field. To show this explicitly, Haldane constructed a theoretical band insulator that provides a minimal model of the integer effect [29]; he considered spinless electrons hopping on a honeycomb lattice, with broken time reversal symmetry but with vanishing average magnetic field, and showed that the Hall conductance could be 0 or ± 1 (in units of e^2/h), depending on parameters.[3] It is quite straightforward to see that such an effect must break $\mathcal{T}$: in the equation $J^x = \sigma_H E^y$, the current J breaks $\mathcal{T}$ (run time backward and the current will flow backward, not stay the same) but the electric field E does not, so the material property σ_H must capture a symmetry breaking in the material.

Note that something quite mysterious must be going on here. These systems are band insulators (the integer states, anyway), whose spectrum includes only "inert" filled bands.

[1] Room temperature in graphene.

[2] In two dimensions, conductance has the units of conductivity.

[3] This sort of thing could also occur in three dimensions, but essentially only as a collection of two dimensional layers.

Apparently these bands are not so inert after all. I noted earlier that this sort of thing could arise in principle: since the Hall current is perpendicular to the applied field, no momentum is transferred to the electrons and no energy dissipated. It may be hard to visualize the mechanism of conduction, since there are no states from which to build wave packets, but at least this phenomenon is consistent with energy conservation.

It turns out to be enlightening to consider the difference between open and periodic boundary conditions. The statement that the Haldane model is a bulk insulator only refers to the bulk of the system and not its boundaries. The standard theoretical device for eliminating the effect of boundaries involves using periodic boundary conditions, but this can mask effects that become apparent with open boundaries, as in cases of topological order. In the present case, it turns out that the edges must be metallic; each edge carries an integer number of "edge modes" (the same integer as the conductance) that are "chiral" (propagate in one direction only) and so capture the broken time reversal symmetry [46, 30]. In a sample with one edge (disk topology), one can then take the view that all the physics occurs at the edge, with the bulk completely inert. This goes by the name "bulk-boundary correspondence".

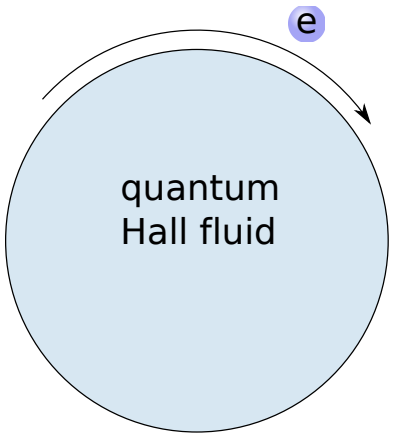

Figure 1.3: The edge of a quantum Hall system carries an electronic mode that is chiral, *i.e.*, that propagates in only one direction.

1.5 Topological insulators with time reversal symmetry

It was thought that there could be no analogous topologically nontrivial band insulators that preserve $\mathcal{T}$. However, a thought experiment by Kane and Mele [39, 40] shows that there should be a two dimensional case. The construction is quite simple. First, suppose there are two kinds of electron (say, in two layers), and the $\mathcal{T}$ breaking for one is equal and opposite to that for the other. Then there is no overall $\mathcal{T}$ breaking, and no net quantum Hall

effect, but there are still chiral edge modes, running in opposite directions. Now suppose that the two kinds of electron are the two values for some component of spin. Spin, as an angular momentum, breaks $\mathcal{T}$, but spin up is precisely the time reverse of spin down, so the two together still preserve $\mathcal{T}$. In particular, the edge mode for one species of spin is the time reverse of the other. It is not too difficult to show that no $\mathcal{T}$-invariant perturbation to the Hamiltonian can mix these states (*i.e.*, no $\mathcal{T}$-invariant operator has a nonzero matrix element between these states), and so no $\mathcal{T}$-invariant mixing of the two species in the bulk of the system will affect the existence of the edge states. The crucial fact is that $\mathcal{T}^2 = -1$ for half-odd spins (spin-1/2 in the case of electrons).

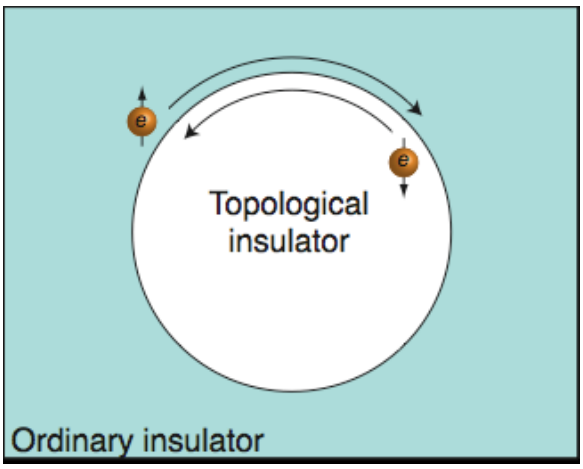

Figure 1.4: The edge of a topological insulator in two dimensions carries a counterpropagating, "helical" pair of electronic modes.

This insensitivity points to the presence of a bulk topological invariant, although it is not given simply by a conductivity or other response function. In this way, the $\mathcal{T}$-invariant topological insulator is a more subtle system than the integer quantum Hall insulator, and probes sensitive to the edge states are needed to verify the presence of the topological state. This state has been reported in HgCdTe/CdTe/HgCdTe wells; the experiment looked for signatures of one-dimensional metallic conduction, characteristic of edge modes [8, 44].

Another difference from the quantum Hall case is the existence of an intrinsically three-dimensional $\mathcal{T}$-invariant topological insulator, not just layered versions of the two-dimensional one just discussed [22, 56, 72]. This system, by contrast, does exhibit a topological bulk response property, in this case the zero-field magnetoelectric polarizability. A number of materials, including BiSb alloy, Bi_2Se_3, and Bi_2Te_3 [37, 85, 14], have been verified to be in this class, although the probes have again been sensitive to the surface states and not to the bulk response function. The surface states of the 3d topological insulator should again span the bulk gap so that the surface is metallic at any chemical potential. In caricature, the surface states look like a cone, which is typically referred to as a "Dirac cone", because a conical 2d dispersion relation can be described by a massless Dirac equation. In particular,

the spin of the surface state at a given momentum is locked to the momentum and rotates through 2π along any curve around the cone. This simple picture is not rigorous since, for example, there need not be a well defined spin state at a given momentum, but there is still a correlation between spin and momentum, and the Berry phase of -1 that results from rotating the spin by 2π is robust.

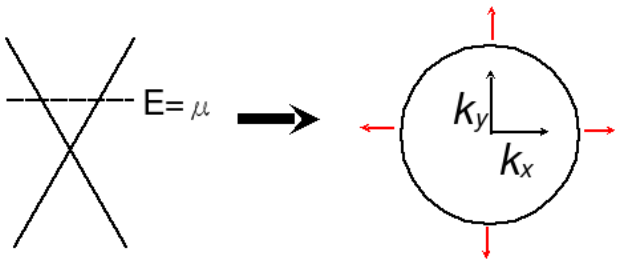

Figure 1.5: The surface of a three-dimensional topological insulators carries an electronic mode with a "massless Dirac" dispersion.

A necessary microscopic property of these materials is the presence of strong spin-orbit coupling; without it, the spin is irrelevant to the electronic structure. Topological insulators are all built of (relatively) heavy elements, which have large nuclear charge Z; since spin-orbit coupling scales as Z^4, it is very large. The doubled Haldane model also crucially involves spin-orbit coupling, since (by construction) its most important constituent is a spin-dependent kinetic term. Finally, notice that the correlation of spin and momentum in the edge modes is crucial for the topological protection seen there.

1.6 Outline

There are (at least) two key questions that need to be asked about the time-reversal-symmetric topological insulators described above. First, how robust are they? That is, can they really be considered to be phases of matter, or are they just the result of some fine-tuning? Since key signatures are seen experimentally, it is clear that there is some robustness, but we need to understand why. In particular, why doesn't the lack of perfect crystal symmetry destroy the states, which were originally predicted in perfect crystals? And why don't weak electronic interactions destroy the states? The work in this dissertation answers these questions, at least in part, by obtaining the topological invariants from a more general perspective than their original formulations.

The second big question which this work addresses is, what physical response properties result from the nontrivial topology? That is, what macroscopic behaviors do the topological insulators display when probed by external fields, in particular electric and magnetic fields? The "Dirac cone" described above can be seen in a kind of microscopy known as ARPES, and can be deduced from scanning tunnelling microscopy as well, but one also wants to know

the macroscopic characteristics as well; this both connects to well-established experimental paradigms in condensed matter physics, and is important if these materials are ever to find practical, technological application. The chapters to follow give partial answers to this crucial question as well.

The remainder of this chapter will introduce in more detail the specific topics to be covered in this dissertation. Chapter 2 reviews some important technical tools and results; in particular, I discuss Bloch's theorem for periodic potentials and the derivation of adiabatic charge transport (the "Kubo formula"). Both these results are crucial for understanding the Hall conductivity as a bulk topological invariant in the integer quantum Hall regime. I have chosen to present the Kubo formula in a many-body formulation, which hopefully provides conceptual clarity, as well as in a relatively idiosyncratic single-particle formulation that explicitly allows degenerate energy levels and that relies on Bloch's theorem (and hence periodicity) minimally.

The original definitions of the topological invariant characterizing the two-dimensional topological insulator rely on the strict periodicity of the material. Chapter 3 reviews one of the more practically useful formulations of the topological invariant and generalizes it to disordered systems. This generalization is important in its own right, and it provides a way to make contact with an old result on 2d systems, namely, that a disordered 2d material with strong spin-orbit coupling should be metallic. The bulk topological invariant does not directly detect this metallic behavior; instead, it reveals the presence of a transition between distinct insulating phases. The transition between integer quantum Hall states remains sharp in the presence of disorder, whereas disorder turns out to broaden the transition between the time-reversal invariant topological insulator and the ordinary insulator. This can be taken as confirmation that a noninsulating, hence metallic, phase intervenes.

Chapters 4 and 5 treat three-dimensional band insulators. In contrast to the two-dimensional case studied in Chapter 3, the three-dimensional topological invariant (or one formulation of it, anyhow) gives a quantized magnetoelectric response [67], very analogous to the quantized Hall conductivity in integer quantum Hall systems. The magnetoelectric response is the response of the electronic polarization $\mathbf{P}$ to an applied magnetic field $\mathbf{B}$, or alternatively the response of the magnetization $\mathbf{M}$ to an applied electric field $\mathbf{E}$. In the absence of dissipation, as in an insulator (so long as the field frequency is smaller than the gap to excitations), these two responses are equal.

The most important result in chapter 4 is that, like the quantum Hall case, this quantized response is (at least in principle) robust to disorder and interactions, and therefore serves to define topological insulators absolutely generally. This result follows directly from now-standard results on the electronic contribution to the polarization of bulk systems, together with some simple geometry. I will review the modern theory of polarization in Chapter 2 as part of the technical introduction, to serve as background for this result. Chapter 4 also uses a nice result from the modern treatment of semiclassical band dynamics [87] to give a simple derivation of the expression for the invariant in the noninteracting case, first derived by evaluating a Feynman diagram in a 4+1 dimensional field theory and then dimensionally reducing down [67]. Finally, in Chapter 4 I show that it is possible to evaluate

the topological invariant in a number of ways, both directly from the derived expression and more physically as the polarization response to an applied magnetic field or as the surface Hall conductivity of a slab. To make these results more general, I have also extended the first model Hamiltonian of a 3d topological insulator [22], breaking time reversal symmetry by adding a coupling to an antiferromagnetic parameter, which allows a smooth interpolation (or "adiabatic continuation") between the topological and trivial phases. This interpolation can almost be seen as an extra dimension in the problem, making contact at some level with the original derivation of the response.

Chapter 5 revisits the general problem of the magnetoelectric response, without much specific reference to topological insulators, and derives the full orbital contribution to the DC linear magnetoelectric response of a crystalline band insulator in the "frozen lattice" limit.[4] From the perspective of computing the properties of real materials, this can be seen as a very limited result; however, it proves to be the most technically challenging part of the linear response problem, and in fact it is already known how to compute all other contributions. Furthermore, the predicted value of this orbital coupling for topological insulators is comparable to measured values of the full magnetoelectric effect in a benchmark material, Cr_2O_3 [32], so the formulae presented here may prove useful as the interest in topological insulators and related compounds continues to increase. Notably, the fully quantum mechanical derivation of Chapter 5 is able to compute explicitly terms that the semiclassical method of Chapter 4 cannot. In particular, Chapter 5 provides a solution to the problem of computing the response to a uniform magnetic field in a crystal, and it is my hope that the methods developed here may prove more generally useful for such problems.

[4]This work was done concurrently with Ivo Souza and Andrei Malashevich at UC Berkeley, who gave a different derivation of the same results and who have published them in a concurrent paper [51].

Chapter 2

Technical review and introduction

2.1 Insulators

The wave function of N electrons moving in a fixed background potential at zero temperature is described by a fully antisymmetric function of the electron coordinates and spins. In the picture provided by Kohn, the defining property of an electrical insulator is that this wave function is composed of localized, nonoverlapping pieces in the $3N$-dimensional configuration space [45].[1] The essential point can be seen by considering a single electron on a ring of length L. Qualitatively, if the wave function wraps all the way around the ring, the state can carry a current, but if the state is localized (decays exponentially with a length scale substantially shorter than the ring, say) the charge cannot move and so there cannot be a current.

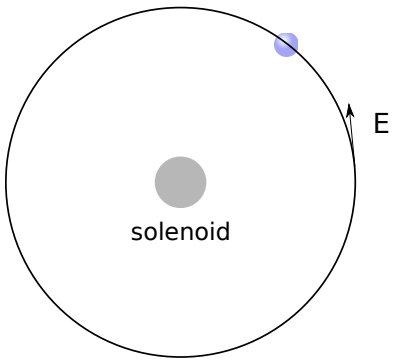

Figure 2.1: A time dependent flux generated by the central solenoid produces an EMF, or electric field, in the ring.

Kohn introduces the key theoretical device of introducing a uniform vector potential A into the system, equivalent to threading a magnetic flux through the ring, assuming the

[1] Really, dN-dimensional, since I also want to discuss one- and two-dimensional insulators.

electron state to be the ground state (or at least an eigenstate) of an appropriate Hamiltonian H. The vector potential enters H with the momentum in the combination $p - eA$, where e is the electron charge; in a tight-binding model, it enters as a complex phase in the hopping parameters, $\exp[(ie/\hbar)\int d\ell A]$. More generally, and more usefully here, A is the conjugate variable to the electron current,[2]

$$\hat{j} = -\frac{\partial H}{\partial A}, \quad j = -\frac{\partial E_0}{\partial A}, \tag{2.1}$$

where E_0 is the energy of the state. This is essentially trying to drive a current adiabatically by slowly ramping up the flux through the ring – recall that A could be interpreted as Et as well as Bx.

The crucial observation is this: if the state is well localized and does not wrap around the ring, the presence of the vector potential can be removed by a gauge tranformation

$$A \to A - \partial_x \chi, \quad \chi = Ax. \tag{2.2}$$

The energy cannot depend on A, then, since a gauge transformation is not physical. On the other hand, if the electron state extends all the way around, then the effect of the gauge transformation on the wave function becomes important: the new wave function

$$\psi(x) \to e^{ieAx/\hbar}\psi(x) \tag{2.3}$$

does not satisfy the right boundary condition, since it is not single-valued (periodic). Hence, this is not a valid gauge transformation when the electron state is extended, and so the state can carry a current. In a strict sense the gauge transformation is not valid for the localized function either, but because there is a point on the ring where the function is vanishingly small, the mismatch is deemed negligible.[3] Finally, notice for later use that when $A = nh/e$ for any integer n, the new wave function *does* satisfy the correct boundary condition, and the effect of the gauge potential really can be removed by a gauge transformation, so that the Hamiltonian is periodic in A with periodicity h/e.

In this way, the problem of currents in a material can be related to the sensitivity of the wave functions to boundary conditions, or alternatively to the topology of the space (here a ring).

For a system with many noninteracting electrons, localization of the many-body wave function is an unnecessarily unwieldy criterion, since all the information in the wave function is encoded in the single particle density matrix $\rho(\mathbf{r}, \mathbf{r}')$, or alternatively in a set of single particle orbitals $\psi(x)$.[4] The density matrix ρ is just the Fermi-Dirac distribution (the temperature will always be assumed to vanish),

$$\rho = \theta(\mu - H), \tag{2.4}$$

[2]Since A is uniform here, there is no need to worry about operator ordering, although it is a legitimate concern in general.

[3]In fact, this argument is not quite sufficient to show that the conductivity vanishes.

[4]Dirac's original discussion of the density matrix [17] is still illuminating.

where θ is the Heaviside step function and μ is a chemical potential. The two crucial properties of this operator for what follows are

- $\rho^2 = \rho$ (idempotency): ρ is a projection operator onto the filled single-electron states.
- $[H, \rho] = 0$: ρ is an eigenoperator of the Hamiltonian.

It is important to realize that all the occupied electronic levels, that is, those with energy less than μ, are degenerate when considered as eigenstates of ρ, and similarly for the unoccupied levels; the former have eigenvalue 1, the latter 0. Physical quantities should not depend on the eigenstates and energies of H individually, but only as a group. This is effectively another kind of "gauge invariance", a redundancy introduced by using the (otherwise more convenient) description of the many-electron wave function in terms of the single particle density matrix.

2.2 Adiabatic conduction, charge pumping, and geometry

It will prove useful to have a more quantitative account of an adiabatically driven current. I will provide two derivations, one in terms of the many electron wave function and one in terms of the one electron density matrix. The formula generated is called a Kubo formula,[5] although zero-temperature derivations of the Kubo formula for conductivity typically assume an electric field that oscillates at a finite frequency. There is another useful formalism, "semiclassical dynamics", which I will not review here but which is the basis for the main derivation in Chapter 4. A comprehensive treatment from a modern viewpoint is given by Xiao, Chang, and Niu [86]. That review also covers much of the material in this section and the next, but it seems worthwhile to include it here anyway.

The adiabatic approximation to the ground state of a Hamiltonian $H(\lambda(t))$ that varies slowly is just the instantaneous ground state ψ_0,

$$\psi(t) \approx e^{i\phi(t)}\psi_0(\lambda), \quad H(\lambda)\psi_0(\lambda) = E_0(\lambda)\psi_0(\lambda), \quad H(\lambda) = H(0) + \lambda\mathcal{O}_\lambda. \tag{2.5}$$

The phase factor takes care of the usual dynamical evolution in a time-independent Hamiltonian, as well as the arbitrariness inherent in identifying the ground states at different values of the parameter. However, this is not a satisfactory approximation here; it predicts that the current driven by an adiabatic change is just

$$\langle\psi_0(\lambda)|\mathbf{j}|\psi_0(\lambda)\rangle, \tag{2.6}$$

which is just the current that would exist if the Hamiltonian were not varying. At the next order of approximation,

$$|\psi(t)\rangle \approx \exp i\phi(t)\left[|\psi_0(\lambda)\rangle + i\hbar\dot{\lambda}\sum_{m\neq 0}|\psi_m(\lambda)\rangle\frac{\langle\psi_m(\lambda)|\partial_\lambda\psi_0(\lambda)\rangle}{E_m(\lambda) - E_0(\lambda)}\right], \tag{2.7}$$

[5] Or Kubo-Greenwood or even Kubo-Greenwood-Peierls

wherein it can be seen explicitly that the accuracy of the approximation is governed by the size of the excitation gap $E_1 - E_0$ in addition to the rate of change $\dot{\lambda}$.[6] Then

$$\langle \mathcal{O}_{\lambda'} \rangle(t) = i\hbar\dot{\lambda} \sum_{m \neq 0} \frac{\langle \psi_0(\lambda) | \mathcal{O}_{\lambda'} | \psi_m(\lambda) \rangle \langle \psi_m(\lambda) | \partial_\lambda \psi_0(\lambda) \rangle}{E_m(\lambda) - E_0(\lambda)} + \text{c.c.}, \tag{2.8}$$

assuming the zeroth order term vanishes. Now, a very useful relation follows directly from the definition of the instantaneous eigenstates,

$$H(\lambda)\psi_0(\lambda) = E_0(\lambda)\psi_0(\lambda) \Rightarrow \langle \psi_m(\lambda) | \partial_\lambda \psi_0(\lambda) \rangle = \frac{\langle \psi_m(\lambda) | \partial_\lambda H(\lambda) | \psi_0(\lambda) \rangle}{E_m(\lambda) - E_0(\lambda)}. \quad (m \neq 0) \tag{2.9}$$

This leads to two alternative formulations:

$$\langle \mathcal{O}_{\lambda'} \rangle(t) = i\hbar\dot{\lambda} \sum_{m \neq 0} \frac{\langle \psi_0 | \mathcal{O}_{\lambda'} | \psi_m \rangle \langle \psi_m | \mathcal{O}_\lambda | \psi_0 \rangle - \text{c.c.}}{(E_m - E_0)^2} \tag{2.10}$$

and

$$\langle \mathcal{O}_{\lambda'} \rangle(t) = i\hbar\dot{\lambda} \left[\langle \partial_{\lambda'} \psi_0 | \partial_\lambda \psi_0 \rangle - \langle \partial_\lambda \psi_0 | \partial_{\lambda'} \psi_0 \rangle \right]. \tag{2.11}$$

There is really a factor $1 - |\psi_0\rangle\langle\psi_0|$ in the latter inner products, but it drops out of the real part. In the case of currents, the conjugate variables λ, λ' are given by components of the vector potential,

$$\langle j^i \rangle(t) = i\hbar E^j \left[\langle \partial_{A_i} \psi_0 | \partial_{A_j} \psi_0 \rangle - \langle \partial_{A_j} \psi_0 | \partial_{A_i} \psi_0 \rangle \right], \tag{2.12}$$

that is,

$$\sigma^{ij} = i\hbar \left[\langle \partial_{A_i} \psi_0 | \partial_{A_j} \psi_0 \rangle - \langle \partial_{A_j} \psi_0 | \partial_{A_i} \psi_0 \rangle \right]. \tag{2.13}$$

Notice that the diagonal conductivity, indeed the symmetric part, vanishes automatically: the adiabatic approximation applies precisely when energy is not being pumped into the system (excitations are gapped), and so there cannot be a longitudinal current (which dissipates energy).

The derived expression for the conductivity has a geometric interpretation. Wave functions Ψ live in a complex projective space, which we usually imagine to be embedded in a Hilbert space. A natural metric exists to measure distances and angles, acting on tangent vectors to the Hilbert space $\delta\Psi$. This is called the Fubini-Study metric, and takes the form

$$g(\delta\Psi_1, \delta\Psi_1)(\Psi) = \langle \delta\Psi_1 | (1 - |\Psi\rangle\langle\Psi|) | \delta\Psi_2 \rangle, \tag{2.14}$$

where Ψ is again a point in the complex projective space, while the inner product is the Hilbert space inner product [94]. The projector $1 - |\Psi\rangle\langle\Psi|$ appears because any component of $\delta\Psi$ parallel to Ψ does not point to a state physically different from Ψ. Now, if $\Psi = \psi_0(\lambda, \lambda')$, $\delta\Psi_1 = \partial_\lambda \psi_0 \delta\lambda$, and $\delta\Psi_2 = \partial_{\lambda'} \psi_0 \delta\lambda'$, this becomes

$$g_{\lambda\lambda'} \delta\lambda \delta\lambda' = \langle \partial_\lambda \psi_0 | (1 - |\psi_0\rangle\langle\psi_0|) | \partial_{\lambda'} \psi_0 \rangle \delta\lambda \delta\lambda', \tag{2.15}$$

[6] For an immensely detailed derivation of this old result, see [16].

and so

$$\sigma^{ij} = -2\hbar \,\mathrm{Im}\left(g_{A_i A_j}\right). \tag{2.16}$$

A more popular geometric interpretation is in terms of a curvature instead of a metric. In this case we write

$$\begin{aligned} f_{\lambda\lambda'} &= \langle \partial_\lambda \psi_0 | \partial_{\lambda'} \psi_0 \rangle - \langle \partial_{\lambda'} \psi_0 | \partial_\lambda \psi_0 \rangle \\ &= \partial_\lambda \langle \psi_0 | \partial_{\lambda'} \psi_0 \rangle - \partial_{\lambda'} \langle \psi_0 | \partial_\lambda \psi_0 \rangle \\ &= \partial_\lambda a_{\lambda'} - \partial_{\lambda'} a_\lambda, \qquad\qquad a_\lambda \equiv \langle \psi_0 | \partial_\lambda \psi_0 \rangle, \end{aligned} \tag{2.17}$$

where a_λ is called the Berry connection and $f_{\lambda\lambda'}$ the Berry curvature. This a really is the connection of the fiber bundle created by projecting the trivial bundle $\boldsymbol{\lambda} \times \mathcal{H}$, with $\boldsymbol{\lambda}$ the parameter space (λ, λ') and $\mathcal{H}$ the Hilbert space, onto the ground-state subspace with $|\psi_0\rangle\langle\psi_0|$, and f is the corresponding curvature. These definitions make a and f imaginary, or more generally anti-Hermitian; later on, it will sometimes be convenient to use a Hermitian definition, $\mathcal{A} = ia$ and $\mathcal{F} = if$.

The connection and curvature are named after Berry, who explored the effects of this adiabatic geometric structure [9]. In particular, he noticed that after an adiabatic variation of parameters that forms a closed loop in $\boldsymbol{\lambda}$, the true ground state wave function picks up a phase determined solely by a (or f) and quite independent of the usual factor $\int E_0(t)dt/\hbar$. Indeed, it is possible to interpret both the Aharonov-Bohm effect and the nontrivial behavior of a spin-1/2 under a 2π rotation as Berry phases.

Thouless [79] pointed out that such an effect can occur with a one-parameter adiabatic variation if the Hamiltonian is periodic in the parameter. In particular, consider the insulator Hamiltonian on a ring, as discussed briefly earlier. Then the parameter λ might be the position of a potential well that could be dragged slowly around the ring, for example. The current driven by any such variation that brings the Hamiltonian back to its original state is, from before,

$$j = i\hbar\dot{\lambda}[\partial_\lambda a_A - \partial_A a_\lambda]. \tag{2.18}$$

The total charge that flows will be given by

$$\Delta Q = i\frac{e}{2\pi}\int_0^1 d\lambda \int_0^{h/e} dA\,[\partial_\lambda a_A - \partial_A a_\lambda], \tag{2.19}$$

where the periodicity of the Hamiltonian in λ is assumed to be 1 and the periodicity in A is h/e as noted earlier. The average over A is an admittedly strange thing to do, but since the system is assumed to be an insulator, Kohn's criterion allows it.

The integral can be argued to be quantized in units of $2\pi i$. The essence of the argument requires seeing the relation between the connection a and the phase of the instantaneous ground state wave function ψ_0. In particular, changes in phase are intimately tied to changes in the former,

$$\psi_0 \to e^{i\chi(A,\lambda)}\psi_0 \Rightarrow a_\lambda \to a_\lambda + i\partial_\lambda \chi, \tag{2.20}$$

for example. Using Stokes' theorem, the charge can be expressed as the line integral of the "vector potential" a around the boundary of the integration region, which by the last relation is tied to the winding of the phase around that boundary. If ψ_0 is to be single-valued as a function of parameters, this winding must be an integer multiple of 2π.

This argument means that $\Delta Q = ne$ for some integer n, that is, that adiabatic charge transport, or pumping, is quantized in units of the fundamental charge. It is therefore our first example of a topological quantity characterizing a physical process.

From a mathematical point of view, the charge pumped is given by the first Chern class of the ground state fiber bundle over the parameter space. As far as the Hamiltonian is concerned, the parameter space (A, λ) is a torus T^2, since it is periodic in both parameters; so is the ground state projector $|\psi_0\rangle\langle\psi_0|$, which generates the bundle in question. The Chern class of a one-dimensional bundle over a compact two-dimensional manifold like T^2 is

$$C_1 = \frac{i}{2\pi}\int d^2\lambda\, f_{\lambda\lambda'}. \tag{2.21}$$

The existence of such an integer can be deduced through a related construction. The above construction computes the "cohomology class" of the fiber bundle, but it is sometimes helpful to think about the "homotopy" classes instead. In this approach, the goal is to find the set of topologically distinct ways that a torus can be embedded into the space of projection matrices.

A simple example will serve to illustrate the connection. Consider a Hamiltonian

$$H(\lambda, \lambda') = \hat{\mathbf{n}}\cdot\boldsymbol{\sigma}, \quad \hat{\mathbf{n}}(\lambda, \lambda') = (\cos\phi\sin\theta, \sin\phi\sin\theta, \cos\theta) \tag{2.22}$$

where the spherical angles ϕ and θ are functions of the parameters. Because $H^2 = 1$, it is easy to see that the ground state projector is

$$\mathcal{P} = \frac{1-H}{2} = \begin{pmatrix} \sin^2\frac{\theta}{2} & -\cos\frac{\theta}{2}\sin\frac{\theta}{2}e^{-i\phi} \\ -\cos\frac{\theta}{2}\sin\frac{\theta}{2}e^{i\phi} & \cos^2\frac{\theta}{2}\end{pmatrix} = \begin{pmatrix}\sin\frac{\theta}{2} & -\cos\frac{\theta}{2}e^{-i\phi}\end{pmatrix}\begin{pmatrix}\sin\frac{\theta}{2} \\ -\cos\frac{\theta}{2}e^{i\phi}\end{pmatrix}, \tag{2.23}$$

and from the ground state spinor it is straightforward to compute[7]

$$f_{\lambda\lambda'} = -\frac{i}{2}\sin\theta\frac{\partial(\theta,\phi)}{\partial(\lambda,\lambda')}. \tag{2.24}$$

Naïvely this makes

$$C_1 = \frac{1}{4\pi}\int \sin\theta d\theta d\phi = 1, \tag{2.25}$$

but this is not quite right. For example, $\partial(\theta,\phi)/\partial(\lambda,\lambda')$ can take both positive and negative values, but the simple result given assumed that it was always positive. More importantly,

[7]For the computation, the differential forms notation simplifies things. In this notation, $\boldsymbol{a} = \langle\psi|\mathbf{d}|\psi\rangle$ and $\boldsymbol{f} = \mathbf{d}\boldsymbol{a}$, where the differential operator is $\mathbf{d} = \sum_i d\lambda^i\partial_{\lambda^i}$ and the differentials anticommute, $d\lambda^i d\lambda^j = -d\lambda^i d\lambda^j$.

the integral over λ, λ' could cover the unit sphere θ, ϕ more than once, in which case there would need to be an integral over the sphere for each covering. This means that C_1 is just the "degree" of the map $\theta(\lambda, \lambda'), \phi(\lambda, \lambda')$. This is the idea of homotopy, which counts the ways that one space can be mapped into another. The simple connection between cohomology and homotopy indicated here does not hold in all cases, but it is important to recognize.

2.3 Electronic polarization (and magnetization)

Maybe it's time to return to a more physical question for a while. In the charge pumping example, what if the parameter λ is not varied through its full range? In that case, the pumped charge need not be an integer. Suppose further that the ring-like system is composed of a finite sample of the insulating material of interest and a voltmeter that connects the two ends. Then the charge pumped is really the charge that flows through the zero-dimensional surface of the material. In the absence of the voltmeter, that charge would build up at the ends; in an insulator, it cannot flow back away from the surface. By allowing it to flow through the voltmeter (which is really just a very large resistor, that measures the charge pumped through the known resistance) this experimental paradigm prevents a static electric field due to the charge buildup from developing. In this way, one can compute an intrinsic difference in the end charges between the initial and final states of the material.

In fact, this is the experimental procedure to determine the polarization of an insulator. This follows from the relation $\sigma = \mathbf{P} \cdot \hat{\mathbf{n}}$, where σ is the surface charge density (in this one-dimensional example, just the end charge), $\mathbf{P}$ is the polarization (dipole moment density), and $\hat{\mathbf{n}}$ the surface normal. Two important points are worth noting: 1. This definition relies on the existence of a reference material with vanishing polarization, to which the substance of interest can be adiabatically deformed (at least theoretically); 2. Since the periodic variation could in principle be run through an arbitrary number of times, there is an arbitrariness to the polarization, with the discrete unit of arbitrariness e.

As a slight digression, the magnetization can be obtained by linear response methods as well, and the result comes out more simply than in the case of polarization. The electromagnetic energy current density is given by the Poynting vector, $\mathbf{S} = \mathbf{E} \times \mathbf{H} = \mathbf{E} \times \mathbf{B}/\mu_0 - \mathbf{E} \times \mathbf{M}$. Therefore, a computation of the response of the energy current $\mathrm{Tr}[\rho H \mathbf{v}]$, or at least a properly symmetrized version thereof, to an applied electric field will yield the zero-field magnetization. The standard derivation of the orbital magnetization [78] starts from a finite system and shows that the result can be written solely in terms of bulk quantities; this alternate approach has the nice feature that it works entirely in the bulk.

2.4 Bloch's theorem

From this point forward, almost all results will assume that electrons do not interact with another except through a mean-field potential. Most results will also assume a crystalline system, in which case the lattice translation symmetry can be used to simplify the problem

of finding the ground state of the system, and thereby its response functions. The classic result is Bloch's theorem: the stationary states $\psi_{n\mathbf{k}}$ of a crystalline Hamiltonian satisfy the following properties:

$$H\psi_{n\mathbf{k}} = E_{n\mathbf{k}}\psi_{n\mathbf{k}}, \qquad \psi_{n\mathbf{k}}(\mathbf{r}+\mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}}\psi_{n\mathbf{k}}(\mathbf{r}). \tag{2.26}$$

The quantum number $\mathbf{k}$, called the wave number or the crystal momentum, takes values in any primitive unit cell of the reciprocal lattice, which is the set of vectors $\mathbf{G}$ such that $\mathbf{G}\cdot\mathbf{R} = 2\pi m$ for any integer m and any "direct" lattice vector $\mathbf{R}$. It is also possible to allow $\mathbf{k}$ to take any value in the reciprocal lattice, in which case

$$\psi_{n\mathbf{k}+\mathbf{G}} = \psi_{n\mathbf{k}}. \tag{2.27}$$

I will use the name "the Brillouin zone" (BZ) for any useful choice of the primitive cell of the reciprocal lattice, even though that term refers to a particular choice of cell. Because of the periodicity in $\mathbf{k}$, the BZ can be viewed as a torus, although one sometimes needs to be careful about this.

The second characteristic property of the wave functions means they can be decomposed into a plane wave part, with wave number inside the BZ, and a periodic part u whose Fourier components have wave numbers in $\{\mathbf{G}\}$,

$$\psi_{n\mathbf{k}} = e^{i\mathbf{k}\cdot\mathbf{r}}u_{n\mathbf{k}}, \qquad u_{n\mathbf{k}}(\mathbf{r}+\mathbf{R}) = u_{n\mathbf{k}}(\mathbf{r}). \tag{2.28}$$

Notice that $u_{n\mathbf{k}+\mathbf{G}} = e^{-i\mathbf{G}\cdot\mathbf{r}}u_{n\mathbf{k}}$ to be consistent with the periodicity of ψ, which is why it is good to be cautious about considering the BZ a torus. It can be useful to think of the relation between u and ψ as a unitary transformation; applying the same transformation to the Hamiltonian allows us to define

$$H_{\mathbf{k}} \equiv e^{-i\mathbf{k}\cdot\mathbf{r}}He^{i\mathbf{k}\cdot\mathbf{r}}, \tag{2.29}$$

which I will call the Bloch Hamiltonian. The Bloch Hamiltonian is periodic in the spatial coordinate $\mathbf{r}$ but, like the functions u it operates on, not in the reciprocal coordinate $\mathbf{k}$.

The crucial object is the ground state density matrix,

$$\rho = \sum_{n\,\mathrm{occ},\mathbf{k}} |\psi_{n\mathbf{k}}\rangle\langle\psi_{n\mathbf{k}}|, \qquad \rho(\mathbf{r},\mathbf{r}') = \sum_{n\,\mathrm{occ},\mathbf{k}} \psi_{n\mathbf{k}}(\mathbf{r})\psi^*_{n\mathbf{k}}(\mathbf{r}'). \tag{2.30}$$

For future reference, note that $\sum_{\mathbf{k}} = \Omega\int_{\mathrm{BZ}}\frac{d^3k}{(2\pi)^3} = N\Omega_0\int_{\mathrm{BZ}}\frac{d^3k}{(2\pi)^3}$, where Ω is the volume of the crystal, with N unit cells of volume Ω_0. It may be instructive to look at the difference between band metals and insulators in terms of the density matrix. For a metal, consider a one dimensional crystal with a band structure symmetric in k:

$$\rho(\mathbf{r},\mathbf{r}') = \frac{Na}{2\pi}\int_{-k_F}^{k_F} dk\, e^{ik(r-r')}u_{nk}(r)u^*_{nk}(r'). \tag{2.31}$$

Ignoring the presence of the u^s for the moment, the integral gives $2\sin[k_F(r-r')]/(r-r')$, where k_F is the Fermi wave vector marking the boundary of the filled subspace. Two points are noteworthy: 1. The decay of the matrix elements is algebraic,[8] so that they are not really short-ranged; 2. The oscillations become more rapid as k_F increases. In the insulating case, when there is no Fermi wave vector, the result is most clear on a lattice, in which case $\rho(R,R') = N\delta_{R,R'}$, and the correlations are short ranged. These qualitative features persist when the factors of u are restored; in the continuum case, the decay will be exponential in an insulator.

2.5 Adiabatic conduction and geometry revisited, with projectors

Now we can derive the charge pumping results using the density matrix to describe the ground state instead of the many-body wave function.

$$J^i(t) = \frac{e}{\Omega}\mathrm{Tr}\rho(t)v^i \tag{2.32}$$

The lowest adiabatic approximation is $J \sim \rho v$ with the instantaneous ground state ρ; this current vanishes by assumption, and we need to know the leading adiabatic correction to the density matrix. Since the leading correction involves

$$\dot{\rho} = \frac{1}{i\hbar}[H,\rho], \tag{2.33}$$

it will be useful to have the Hamiltonian appear explicitly, as in

$$\mathbf{v} = \frac{1}{i\hbar}[\mathbf{r},H]. \tag{2.34}$$

Note that, while this looks like Heisenberg's equation of motion, all equations are actually in the Schrödinger representation, and this relationship is just the appropriate generalization of the more familiar $[r,.] = i\hbar\partial_p$. In any case,

$$J^i(t) = \frac{ie}{\hbar\Omega}\mathrm{Tr}\rho(t)[H,r^i] = \frac{ie}{\hbar\Omega}\mathrm{Tr}H[r^i,\rho(t)]. \tag{2.35}$$

The last equality follows from cyclicity of the trace, but it is important to recognize that not all expressions involving the operator r are admissible. In particular, the form given looks like

$$\int d\mathbf{r}_1 d\mathbf{r}_2 H_{12}(r_2^i - r_1^i)\rho_{21} \tag{2.36}$$

in the position basis. The term $(r_2^i - r_1^i)$ is potentially problematic because it diverges in some directions in the six-dimensional configuration space, but this problem is avoided because

[8] That is, power-law rather than exponential.

ρ_{21} is exponentially suppressed in $\mathbf{r}_2 - \mathbf{r}_1$. On the other hand, the alternate form $r^i[\rho, H]$ looks like $\int d\mathbf{r}_1 r_1^i [\rho, H]_{11}$, which diverges in $\mathbf{r}_1$, and so is not really admissible.

To proceed, we need to know an important property of projection operators. Consider a projector $\mathcal{P}$ and its complement $\mathcal{Q}$, which will be functions of some parameter λ. It is straightforward to show that

$$\mathcal{P}(\partial_\lambda \mathcal{P})\mathcal{P} = \mathcal{Q}(\partial_\lambda \mathcal{P})\mathcal{Q} = 0, \tag{2.37}$$

that is, that the first order variation or derivative of a projection operator has no "interior" matrix elements. For the density matrix, this means that matrix elements between pairs of occupied states, or pairs of unoccupied states, vanish. In fact, this property extends beyond derivatives, to any "derivation", or operation that satisfies the product rule. Of most importance for our current purpose, a commutator is a derivation, and so it follows that

$$[r^i, \rho(t)] = \rho(t)[r^i, \rho(t)] + [r^i, \rho(t)]\rho(t), \tag{2.38}$$

and, with some algebra that makes use of idempotency ($\rho^2 = \rho$), cyclicity, and the above considerations,

$$\begin{aligned} J^i(t) &= \frac{ie}{\hbar\Omega}\mathrm{Tr}\rho(t)[r^i, \rho(t)][H, \rho(t)] - \text{c.c.} \\ &= \dot{\lambda}\frac{e}{\Omega}\mathrm{Tr}\rho(t)[\rho(t), r^i]\partial_\lambda \rho(t) + \text{c.c.} \end{aligned} \tag{2.39}$$

This expression is first order in the small parameter $\dot{\lambda}$, so this is the right point in this argument to take the adiabatic approximation and drop the dependence on t,

$$J^i = \dot{\lambda}\frac{e}{\Omega}\mathrm{Tr}\rho[\rho, r^i]\partial_\lambda \rho + \text{c.c.} \tag{2.40}$$

To find the quantum Hall response, we need to choose $\lambda = -A_j$ again. We also need to know how to write perturbation theory with density matrices. For the instantaneous ground state,

$$[\rho, H] = 0 \Rightarrow [\partial_\lambda \rho, H] = [\partial_\lambda H, \rho] = e[v^j, \rho] = \frac{e}{i\hbar}[[r^j, H], \rho] = -\frac{e}{i\hbar}[[\rho, r^j], H]. \tag{2.41}$$

The last step used the Jacobi identity and $[\rho, H] = 0$. This equation essentially defines $\partial_\lambda \rho$, and it is acceptable to take

$$\partial_\lambda \rho = -\frac{e}{i\hbar}[\rho, r^j]. \tag{2.42}$$

Then

$$\sigma^{ij} = \frac{i}{\Omega}\frac{e^2}{\hbar}\mathrm{Tr}\rho[\rho, r^i][\rho, r^j] - \text{c.c.} \tag{2.43}$$

That expression is quite general. In fact, it expresses a sort of noncommutative geometry defined by the fermionic ground state. The coordinate of this geometry is

$$\boldsymbol{z} = i[\rho, \mathbf{r}], \tag{2.44}$$

and

$$\sigma^{ij} = \frac{e^2}{i\hbar}\frac{1}{\Omega}\mathrm{tr}[\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j], \tag{2.45}$$

where tr in lower case is taken in the occupied subspace only. This looks just like the situation of Landau levels (in a magnetic field) in two dimensions, for which the guiding center coordinates satisfy $[X, Y] = i/2\pi n_\phi$ and $\sigma^{xy} = -(e^2/h)(n_e/n_\phi)$, where n_e is the electronic density and n_ϕ the magnetic flux density. The filling factor n_e/n_ϕ is an integer when Landau levels are filled. It can be seen that $[\hat{\boldsymbol{v}}^i, \hat{\boldsymbol{v}}^j]$ defines a measure of area intrinsic to the quantum state; in the Landau level problem, this area is just $1/n_\phi$.

When the system has crystal symmetry, the density matrix decomposes into sectors of different $\mathbf{k}$. We can write

$$\rho = \int_{\mathrm{BZ}} \frac{d^d k}{(2\pi)^d} e^{i\mathbf{k}\cdot\mathbf{r}} \mathcal{P}_{\mathbf{k}} e^{-i\mathbf{k}\cdot\mathbf{r}}, \qquad \mathcal{P}_{\mathbf{k}} = \sum_{n\,\mathrm{occ}} |u_{n\mathbf{k}}\rangle\langle u_{n\mathbf{k}}|. \tag{2.46}$$

Then

$$\hat{\boldsymbol{v}}^i = i[\rho, r^i] = \int_{\mathrm{BZ}} \frac{d^d k}{(2\pi)^d} e^{i\mathbf{k}\cdot\mathbf{r}} (\partial_{k_i}\mathcal{P}_{\mathbf{k}}) e^{-i\mathbf{k}\cdot\mathbf{r}}, \qquad \hat{\boldsymbol{v}}^i_{\mathbf{k}} = \partial_{k_i}\mathcal{P}_{\mathbf{k}} \tag{2.47}$$

using the periodicity of $\psi_{n\mathbf{k}}$ in $\mathbf{k}$. In two dimensions, this gives

$$\sigma^{ij} = \frac{e^2}{h}\frac{1}{2\pi i}\int_{\mathrm{BZ}} d^2k\, \mathrm{Tr}\mathcal{P}_{\mathbf{k}}[\partial^i\mathcal{P}_{\mathbf{k}}, \partial^j\mathcal{P}_{\mathbf{k}}] \tag{2.48}$$

or

$$\sigma^{ij} = \frac{e^2}{h}\frac{1}{2\pi}\int_{\mathrm{BZ}} d^2k \sum_{n\,\mathrm{occ}} f^{ij}_{nn}, \quad f^{ij}_{nn'}(\mathbf{k}) = \langle\partial_{k_i}u_{n\mathbf{k}}|\mathcal{Q}_{\mathbf{k}}|\partial_{k_j}u_{n'\mathbf{k}}\rangle - (i \leftrightarrow j), \quad \mathcal{Q}_{\mathbf{k}} = \mathbb{1}_{\mathbf{k}} - \mathcal{P}_{\mathbf{k}}. \tag{2.49}$$

The tensor f defined here is the nonabelian (*i.e.*, multicomponent) version of the Berry curvature defined earlier.

The integral of the curvature is known as a Chern invariant, and is quantized in integer steps, as discussed earlier. Analogous invariants appear in all even dimensions. In four dimensions,

$$\frac{1}{2!}\frac{1}{2^2}\frac{1}{(2\pi)^2}\int_{\mathrm{BZ}} d^4k\, \epsilon_{abcd}\mathrm{tr} f^{ab} f^{cd} \tag{2.50}$$

takes integer values, and the matrix product and trace take place solely in the occupied subspace.

The curvature satisfies an equation analogous to the source-free Maxwell equations for electrodynamics. Locally (in $\mathbf{k}$, in this context), this means it can be written as

$$f^{ij}_{nn'} = \partial^i a^j_{nn'} - \partial^j a^i_{nn'} + [a^i, a^j]_{nn'}, \tag{2.51}$$

and the connection a can be taken to be the nonabelian Berry connection

$$a^i_{nn'} = \langle u_{n\mathbf{k}}|\partial^i u_{n'\mathbf{k}}\rangle. \tag{2.52}$$

This allows us to define a further object called a Chern-Simons (CS) form, which lives naturally in odd dimensions. In one dimension, the CS form is just the connection, while in higher dimensions it takes a more complicated form:

$$\begin{aligned} \mathrm{CS1}_i &= \epsilon_{ij}\,\mathrm{tr}\,a^j, \quad \frac{1}{2}\epsilon_{ij}\,\mathrm{tr} f^{ij} = \partial^i \mathrm{tr}\,\mathrm{CS1}_i \\ \mathrm{CS3}_i &= \epsilon_{ijkl}\,\mathrm{tr}\left[a^j\partial^k a^l + \frac{2}{3}a^j a^k a^l\right], \quad \frac{1}{2^2}\epsilon_{ijkl}\,\mathrm{tr} f^{ij} f^{kl} = \partial^i \mathrm{tr}\,\mathrm{CS3}_i. \end{aligned} \tag{2.53}$$

The forms written here[9] require an even-dimensional space (because of the number of indices on the Levi-Civita tensor ϵ), but each component corresponds to an odd-dimensional quantity.

[9] All of these expressions can be cast in the language of the exterior algebra of differential forms, but since they emerge physically from tensors that are not obviously antisymmetric to begin with, I find it more consistent to stick with the tensor notation. However, once an expression is in a form that can be translated into differential forms, that translation can make the algebra significantly easier, especially if you're not worried about factors of 2 and the like.

Chapter 3

Topological insulators with disorder and a spin-orbit metal

3.1 Introduction

Considerable theoretical and experimental effort has been devoted to the quest for an intrinsic spin Hall effect [59, 76, 41, 84] that would allow generation of spin currents by an applied electric field. Interesting mechanisms for such spin current generation make use of spin-orbit coupling, which breaks the $SU(2)$ spin symmetry of free electrons but not time-reversal symmetry. A dissipationless type of intrinsic spin Hall effect was predicted [39, 40] to arise in materials that have an electronic energy gap. This "quantum spin Hall effect" (QSHE) in certain materials with time-reversal symmetry has a subtle relationship to the integer quantum Hall effect, in which time-reversal symmetry is explicitly broken by a magnetic field.

In a system with unbroken time-reversal symmetry, a dissipationless charge current is forbidden, but a dissipationless transverse spin current is allowed, of the form

$$\mathcal{J}^i_j = \alpha \epsilon_{ijk} E_k. \tag{3.1}$$

The current on the left is a spin current and ϵ is the fully antisymmetric tensor. Note that a spin current requires two indices, one for the direction of the current and one for the direction of angular momentum that is transported. The constant of proportionality α depends on the specific mechanism: for example, the (dissipative) extrinsic D'yakonov-Perel mechanism [18] predicts a small α that depends on impurity concentration. The QSHE builds on the construction by Haldane [29] of a lattice "Chern insulator" model, with broken time-reversal symmetry but without net magnetic flux, that shows a $\nu = 1$ IQHE. The simplest example of a QSHE is obtained by taking two copies of Haldane's model, one for spin-up electrons along some axis and one for spin-down. Time-reversal symmetry can be maintained if the effective IQHE magnetic fields are opposite for the two spin components. Then an applied electric field generates a transverse current in one direction for spin-up

electrons, and in the opposite direction for spin-down electrons. There is no net charge current, consistent with time-reversal symmetry, but there is a net spin current. However, models like this in which one component of spin is perfectly conserved are both unphysical, since realistic spin-orbit coupling does not conserve any component, and not very novel, since for each spin component the physics is exactly the same as Haldane's model and the spin components do not mix.

More subtle physics emerges when one asks how the QSHE appears in more realistic band structures. Remarkably, band insulators of noninteracting two-dimensional electrons with spin-orbit coupling divide into two classes: the "ordinary insulator", which in general has no propagating edge modes and no spin Hall effect, and the "topological insulator", which has stable propagating edge modes and a generic spin Hall effect, although the amount of spin transported (the coefficient α in (3.1)) is nonuniversal. These phases are associated with a $\mathbb{Z}_2$-valued topological invariant (an "oddness" or "evenness", in the language of parity[1]) in the same way that IQHE phases are associated with an integer-valued topological invariant. For explicitness, consider the model of graphene introduced by Kane and Mele [39]. This is a tight-binding model for independent electrons on the honeycomb lattice (Fig. 3.1). The spin-independent part of the Hamiltonian consists of a nearest-neighbor hopping, which alone would give a semimetallic spectrum with Dirac nodes at certain points in the 2D Brillouin zone, plus a staggered sublattice potential whose effect is to introduce a gap:

$$H_0 = t \sum_{\langle ij \rangle \sigma} c^\dagger_{i\sigma} c_{j\sigma} + \lambda_v \sum_{i\sigma} \xi_i c^\dagger_{i\sigma} c_{i\sigma}. \tag{3.2}$$

Here $\langle ij \rangle$ denotes nearest-neighbor pairs of sites, σ is a spin index, ξ_i alternates sign between sublattices of the honeycomb, and t and λ_v are parameters.

The insulator created by increasing λ_v is an unremarkable band insulator. However, the symmetries of graphene also permit an "intrinsic" spin-orbit coupling of the form

$$H_{SO} = i\lambda_{SO} \sum_{\langle\langle ij \rangle\rangle \sigma_1 \sigma_2} \nu_{ij} c^\dagger_{i\sigma_1} s^z_{\sigma_1\sigma_2} c_{j\sigma_2}. \tag{3.3}$$

Here $\nu_{ij} = (2/\sqrt{3}) \hat{\boldsymbol{d}}_1 \times \hat{\boldsymbol{d}}_2 = \pm 1$, where i and j are next-nearest-neighbors and $\hat{\boldsymbol{d}}_1$ and $\hat{\boldsymbol{d}}_2$ are unit vectors along the two bonds that connect i to j. Including this type of spin-orbit coupling alone would not be a realistic model. For example, the Hamiltonian $H_0 + H_{SO}$ conserves s^z, the distinguished component of electron spin, and reduces for fixed spin (up or down) to Haldane's model [29]. Generic spin-orbit coupling in solids should not conserve any component of electron spin. However, the unusual phase generated when H_{SO} is strong turns out to survive, with subtle changes, once the spin-orbit coupling is made more realistic, as we now review.

The "topological insulator" phase created when $|\lambda_{SO}| \gg |\lambda_v|$ is quite different from the ordinary insulator that appears when $|\lambda_v| \gg |\lambda_{SO}|$ (here we assume that there is an energy

[1] Throughout "parity" is used to denote oddness or evenness, rather than a spatial inversion eigenvalue.

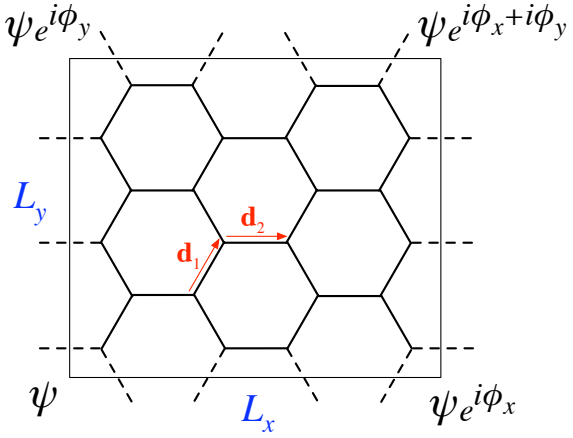

Figure 3.1: The honeycomb lattice on which the tight-binding Hamiltonian resides. For the two sites depicted, the factor ν_{ij} of equation (3.3) is $\nu_{ij} = -1$. The phases $\phi_{x,y}$ describe twisted boundary conditions, introduced in equation (3.10).

gap between the lower and upper band pairs in which the Fermi level lies). The former has counterpropagating edge modes and shows the QSHE, while the latter does not [39]. Does this phase exist for more realistic spin-orbit coupling? The spin component s_z is no longer a good quantum number when the Rashba spin-orbit coupling is added:

$$H_R = i\lambda_R \sum_{\langle ij\rangle\sigma_1\sigma_2} c^\dagger_{i\sigma_1} \left(\boldsymbol{s}_{\sigma_1\sigma_2} \times \hat{\boldsymbol{d}}_{ij}\right)_z c_{j\sigma_2}, \tag{3.4}$$

with $\boldsymbol{d}_{ij}$ the vector from $i \to j$ and $\hat{\boldsymbol{d}}_{ij}$ the corresponding unit vector. Note that Rashba spin-orbit coupling is not intrinsic to graphene but generated by inversion-symmetry breaking in the out-of-plane direction [11]. The Rashba coupling is a standard form that is believed to be a reasonable model for the dominant spin-orbit coupling in adsorbed graphene.

The topological insulator survives but is strongly modified in the presence of the Rashba term. For a general 2D band structure with s^z conserved, there are many phases labeled by an integer n, as in the IQHE: if spin-up electrons are in the $\nu = n$ state, then spin-down electrons must be in the $\nu = -n$ state by time-reversal symmetry, where the sign indicates that the direction of the effective magnetic field is reversed. Once s^z is not conserved, as when $\lambda_R \neq 0$, there are only two insulating phases, the "ordinary" and "topological" insulators. A heuristic definition of the topological insulator, without reference to any particular spin component or the spin Hall effect, is as a band insulator that is required to have gapless propagating edge modes at the sample boundaries. The decoupled $\nu = \pm n$ cases with s^z conserved are adiabatically connected, once s^z is not conserved, to the ordinary insulator for even n and to the topological insulator for odd n. A review of how these two cases emerge as the only possibilities in 2D follows in Section II.

It should be clarified that the intrinsic spin-orbit coupling is now believed to be quite weak in graphene [38, 54, 92], so that the topological insulator is unlikely to be realized. However, the same topological insulator phase is now believed to exist for realistic spin-orbit coupling in other materials such as HgTe [8]. We choose to study the graphene model introduced by Kane and Mele because it is the first and simplest model showing a transition between ordinary and topological insulators. It is the simplest possible model in that it has four spin-split bands, which is the minimum number required for the nontrivial phase to exist [56]. For this reason, it has received the most attention in the other studies [75, 64] to which our results will be compared. It is straightforward to generalize the approach presented here for the graphene model to another material with more complicated spin-orbit coupling, and the same qualitative results are expected to apply.

It is not obvious at first glance how to generalize the topological insulator phase to finite, noncrystalline systems, rather than band structures, as when the parameters of the Hamiltonian $H = H_0 + H_{SO} + H_R$ are drawn from a random distribution. The first approach was in terms of a spin Chern number [75] similar to the Chern integer in finite IQHE systems, but there is now agreement that for a clean band structure the only invariants are of $\mathbb{Z}_2$ type, rather than integer type [23, 56, 26]. Two equivalent definitions of the appropriate $\mathbb{Z}_2$ invariant for a finite disordered system, in the simple case when the disorder splits all degeneracies other than Kramers degeneracies, are as follows (the full definition is given and compared to previous work in the following section). The finite system can be considered as a unit "supercell" of a large 2D lattice. A large, finite supercell gives many bands, but each pair of bands connected by time reversal (Kramers pair) can be assigned its own $\mathbb{Z}_2$ invariant [56]. The phase of the supercell system, if the Fermi level lies in a gap, is then identified by adding up all the invariants (mod 2). Alternately, a direct definition of the phase in the finite system can be given that is related to the notion of "$\mathbb{Z}_2$ pumping" [23]. Real charge is pumped as the flux through the system is taken from 0 to $hc/2e$ (*half* the usual flux quantum that appears in IQHE pumping); we show that in the topological insulator, any pumping cycle, properly defined, pumps an odd number of electron charges, while for the ordinary insulator any cycle pumps an even number of charges.

We implement this definition numerically using an explicit algorithm introduced by Fukui and Hatsugai [26, 25] for computing $\mathbb{Z}_2$ topological invariants on a Brillouin zone. The topological insulator phase is robust to disorder: while different realizations of disorder assign different "Chern parities" to individual subbands, it is found that the total for occupied subbands is always "odd" for a wide range of parameters, which in our definition indicates a topological insulator. In the IQHE, a pair of bands of opposite Chern number can annihilate as the strength of disorder is increased; in the QSHE, two band pairs that both have odd Chern parity can annihilate, i.e., become two even-parity band pairs. If the topological insulator can be destroyed by band annihilation, then there are extended (i.e., topologically nontrivial) states with an arbitrarily small gap; it may be the case that for some range of parameters, there are extended states at the Fermi level even in the thermodynamic limit, indicating a metallic phase.

In the IQHE, there is only a single energy with extended states rather than a range of

energies, and hence no metallic phase. We find the phase diagram of the graphene model with on-site disorder, and in the presence of non-zero Rashba coupling find evidence for a metallic phase intervening between ordinary and topological insulators. The existence of the metallic phase can be understood from work on 2D localization in the symplectic universality class [34], in which time-reversal symmetry is unbroken but spin-orbit coupling is present. It is found that extended states can be stable against this disorder over a nonzero range in energy, unlike in the orthogonal class, in which there are no extended states, or the unitary (integer quantum hall) class, in which there are extended states only at isolated energies. The argument for the existence of the symplectic metal is beyond the scope of this work, but there is a simple picture that captures some of the important ideas.

Consider the question, what is the probability that a spinless electron will return to its starting point? In a path-integral picture, the amplitude associated with some return path γ will be A_γ. In the presence of disorder, this amplitude will interfere incoherently with those from other paths, but it will interfere coherently with the time reversed path (*i.e.*, the same path traversed in the opposite direction), no matter how strong the disorder is. This gives an unnormalized probability $|A + A|^2 = 4|A|^2$. Hence the probability of return is enhanced relative to the the probability to travel to the edge of the system and be lost, since there is no time-reversed path which with to interfere in that case. [2, 1] The situation is different with strong spin-orbit coupling. If the spin is locked to the direction of motion, then there is a spin Berry phase that needs to be included with the amplitudes. In a remarkably simple and comprehensible paper, Bergmann showed that this factor is $-1/2$ on average, due essentially to the fact that a spin-1/2 electron picks up a sign upon rotation through 2π. [7] That is, the interference is destructive rather than constructive, $|A - A/2|^2 = |A|^2/4$, and the spinful electron is more likely to pass through the sample than return to its starting point, leading to a nonzero conductivity.

Our result that there is a metallic phase can be understood as indicating that, although the $\mathbb{Z}_2$ topological invariant allows one to distinguish two kinds of insulators when the Fermi level has no extended states, this invariant does not modify the standard picture of bands of extended states in the symplectic universality class.

Recent work by Obuse *et al.* [63] obtains a phase diagram and critical exponents using a network model for the spin quantum Hall effect that is similar to the Chalker-Coddington network model [13] for the IQHE (see also Onoda *et al.* [64] for a quasi-1D study of localization in the Kane-Mele Hamiltonian with disorder). Our results on the phase diagram are consistent with these works, although our method is unable to generate large enough system sizes to confirm the exponents found for the phase transitions. To understand how the network and Chern-parity approaches complement each other, consider the integer quantum Hall effect (IQHE): while the phenomenological network approach to the IQHE is valuable both to find the critical indices precisely and to identify the minimal necessary elements of a theory for the transition, Chern-number studies remain important for studies of effects such as the floating of extended states [47, 42, 90], where knowledge of the topological properties of a state is required. The network model gives more accurate information about the phase transitions but, if only the localization length is probed, does not distinguish the different

phases in bulk. A more technical difference between the two approaches is discussed at the end of Section II.

Section II reviews how the topological insulator phase in perfect crystals arises from a parity-valued topological invariant of the band structure, similar to the TKNN [81] integers in the IQHE. It then gives two mathematically equivalent definitions of the topological insulator phase in disordered systems based on Chern parity. One definition simply considers a finite disordered system as a supercell of an infinite lattice system, while the other is based upon closed charge pumping cycles driven by application of flux to a finite periodic system. Section III reviews the Fukui-Hatsugai algorithm [26, 25] adapted to numerical computation of these invariants in disordered systems, then computes the phase diagram of the Kane-Mele graphene model [39] with on-site disorder. The conclusions of our study for general 2D disordered systems are summarized in a short Section IV. While there is a three-dimensional version of the QSHE [56, 72, 22] that is less directly connected to the IQHE and has interesting localization behavior, we will restrict our attention to 2D except for some comments in the final section.

3.2 Chern parities for disordered noninteracting electron systems

3.2.1 The definition of the $\mathbb{Z}_2$ invariant in clean systems

We review one definition of the $\mathbb{Z}_2$ invariant of a band pair in a 2D band structure, then explain its generalization to noncrystalline systems. With the Hamiltonian

$$H = H_0 + H_{SO} + H_R \tag{3.5}$$

defined in equations (3.2), (3.3), and (3.4), or any periodic, single-electron Hamiltonian, a Berry connection $\boldsymbol{\mathcal{A}}$ can be defined on the Brillouin zone (BZ) from the periodic part $u(\boldsymbol{k})$ of a Bloch state $\psi_{\boldsymbol{k}} = u(\boldsymbol{k})e^{i\boldsymbol{k}\cdot\boldsymbol{r}}$. For a single nondegenerate band j, $\boldsymbol{\mathcal{A}}_j(\boldsymbol{k}) = -i\langle u_j|\boldsymbol{\nabla}_{\boldsymbol{k}}|u_j\rangle$. This Berry connection serves as a potential for the Berry field strength $\mathcal{F} = (\boldsymbol{\nabla}_{\boldsymbol{k}}\times\boldsymbol{\mathcal{A}})_z$, which, when integrated over the BZ, returns an integer, called a TKNN integer in the context of the IQHE [81]:

$$\begin{aligned} n_j &= -\frac{1}{2\pi}\int_{BZ} \mathcal{F}_j\, d^2k \\ &= \frac{i}{2\pi}\int_{BZ}\left[\left\langle\frac{\partial u}{\partial k_x}\middle|\frac{\partial u}{\partial k_y}\right\rangle - \left\langle\frac{\partial u}{\partial k_y}\middle|\frac{\partial u}{\partial k_x}\right\rangle\right]. \end{aligned} \tag{3.6}$$

When bands touch, only the total TKNN integer of the bands is well-defined [4], in which case it makes sense to generalize $\boldsymbol{\mathcal{A}}$ as

$$\begin{aligned} \boldsymbol{\mathcal{A}}_j &= -i\left(\langle u_{j1}|\boldsymbol{\nabla}_{\boldsymbol{k}}|u_{j1}\rangle + \langle u_{j2}|\boldsymbol{\nabla}_{\boldsymbol{k}}|u_{j2}\rangle\right) \\ &= -i\,\mathrm{tr}\,\boldsymbol{u}_j^\dagger\boldsymbol{\nabla}_{\boldsymbol{k}}\boldsymbol{u}_j \end{aligned} \tag{3.7}$$

for the degenerate Bloch functions u_{j1} and u_{j2}. In the second equation, $\boldsymbol{u}_j = (u_{j1}, u_{j2})$ where the Bloch functions are viewed as column vectors, i.e., $\boldsymbol{u}_j$ is a matrix. This compact notation follows Fukui and Hatsugai [26]. This generalization is always needed in the case of a time-reversal-invariant system with half-odd integer spin, or fermionic statistics, because such a system has "Kramers degeneracies" at certain time-reversal-invariant momenta (see, for example, Ref. [73]). Briefly, time-reversal invariance requires that

$$\Theta \mathcal{H}(-\boldsymbol{k}) \Theta^{-1} = \mathcal{H}(\boldsymbol{k}), \tag{3.8}$$

where $\mathcal{H}(\boldsymbol{k})$ is the Bloch Hamiltonian and $\Theta = -i\sigma_y K$ is the action of time reversal (K performs complex conjugation and σ_y, the usual Pauli matrix, acts on spin indices). At momenta for which $-\boldsymbol{k} = \boldsymbol{k}$, then, $\mathcal{H}(\boldsymbol{k})$ commutes with Θ, but there are no single-electron eigenstates of Θ (time reversal will always flip the spin), so there must be a degenerate pair of energy eigenstates at such momenta. Once we take these degeneracies into account, however, the TKNN integers for the band pairs vanish in a time-reversal-invariant band structure (see, e.g., Ref. [56]). Instead there is a $\mathbb{Z}_2$ invariant associated with a band *pair* in a time-reversal-invariant 2D Fermi system [40].

Fu and Kane [23] give the following formula for the $\mathbb{Z}_2$ topological invariant in terms of the Bloch functions of the clean system:

$$D = \frac{1}{2\pi} \left[\oint_{\partial(EBZ)} d\boldsymbol{k} \cdot \boldsymbol{\mathcal{A}} - \int_{EBZ} d^2k \, \mathcal{F} \right] \mod 2. \tag{3.9}$$

The notation EBZ stands for Effective Brillouin Zone [56], which describes one half of the Brillouin zone together with appropriate boundary conditions. Since the BZ is a torus, the EBZ can be viewed as a cylinder, and its boundary ∂(EBZ) as two circles, as in Fig. 3.2(b). While $\mathcal{F}$ is gauge-invariant, $\boldsymbol{\mathcal{A}}$ is not, and different (time-reversal-invariant) gauges can change the boundary integral by an even amount.

A proof of the existence and $\mathbb{Z}_2$ nature of the topological invariant for multiple bands and some intuition for (3.9) can be obtained [56] by considering time-reversal-invariant band structures as maps from the EBZ to the space of Bloch Hamiltonians, following work on the IQHE by Avron *et al.* [4] The Berry field strength can be written in terms of the (gauge-invariant) projection operator onto the band pair rather than the (gauge-dependent) wavefunctions. In this approach, the ambiguity by an even integer that is crucial to obtain a $\mathbb{Z}_2$ rather than $\mathbb{Z}$ invariant corresponds to the many different ways in which the circles that form EBZ boundaries in Fig. 3.2(b) can be contracted to make the EBZ into a sphere. The boundary integrals in (3.9) just calculate the contribution to the Chern number from these "contractions." On this sphere, Chern integers are well-defined for each nondegenerate band pair, but the different ways of contracting the boundaries cause the resulting integers to differ by even numbers. An explicit numerical implementation of the Fu-Kane formula (3.9) was given by Fukui and Hatsugai [26] and will be reviewed in Section III.

What happens when disorder is added? We now explain how the definition of the topological insulator generalizes to disordered systems. Just as TKNN integers of a band structure

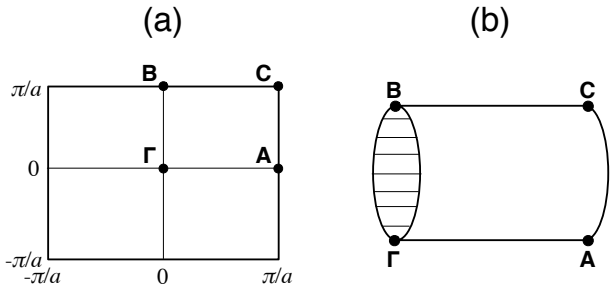

Figure 3.2: (a) A two-dimensional Brillouin zone; note that any such Brillouin zone, including that for graphene, can be smoothly deformed to a torus. The labeled points are time-reversal-invariant momenta. (b) The effective Brillouin zone (EBZ). The horizontal lines on the boundary circles ∂(EBZ) connect time-reversal-conjugate points, where the Hamiltonians are related by time reversal and so cannot be specified independently.

give rise to Chern integers in a disordered system of finite size, the $\mathbb{Z}_2$ invariants of band pairs become "Chern parities".

3.2.2 The topological insulator phase for Slater determinants via Chern parity

The TKNN integers generalize in the presence of disorder and interactions to the Chern number of the many-particle wavefunction [62]. A natural question is how disorder and interactions modify the $\mathbb{Z}_2$ invariants of band pairs in spin-orbit-coupled 2D band structures. All derivations of the $\mathbb{Z}_2$ invariants of clean systems depend on Fermi statistics in some way: for example, the existence of Kramers degeneracies and the related fact that the time-reversal operator squares to -1 both depend on Fermi statistics. A many-fermion wavefunction describing an even number of fermions does not behave in the same way as single-fermion wavefunctions under time-reversal. Hence, given only the many-fermion wavefunction, it does not seem likely that there is a generalization of the $\mathbb{Z}_2$ invariant.

However, for the particular case of many-fermion wavefunctions that are single Slater determinants of single-particle wavefunctions, the invariant can be generalized as we now show. While the assumption of a single Slater determinant limits the treatment of interactions to the Hartree-Fock level, there is no requirement that the wavefunctions in the Slater determinant be Bloch states. As a result, the topological insulator and QSHE can be defined and studied for any disorder strength.

Niu *et al.* [62] and Avron and Seiler [5] showed that for disordered quantum Hall systems there exists a generalization of the TKNN invariant defined for clean systems [81]. They introduce generalized periodic boundary conditions and find an invariant Chern number, similar in form to the TKNN invariant, on the space of boundary phases. Consider a finite system of noninteracting electrons with boundary conditions that are periodic up to phases

ϕ_x, ϕ_y, as shown in Fig. 3.1: this is equivalent to putting magnetic fluxes $\Phi_{x,y} = \phi_{x,y}\Phi_0/2\pi$ through the two noncontractible circles on the torus ($\Phi_0 = hc/e$ is the magnetic flux quantum). As motivation, think of the finite system as a (possibly very large) unit cell of a lattice system. Then in order to determine the phase of this lattice system, instead of integrating over k to do the integrals in the Fu-Kane formula (3.9), we integrate over the boundary phases, which introduce offsets to the wave vectors. We now carry out this procedure, show that it reproduces the band-structure result for clean systems, and then discuss a physical picture and its relation to previous definitions.

Consider the single-particle wavefunctions of a lattice Hamiltonian such as the graphene model on a finite lattice of size $L_x \times L_y$ (see Fig. 3.1). Instead of the physical boundary conditions $\psi(\boldsymbol{x} + \boldsymbol{L}_{x,y}) = \psi(\boldsymbol{x})$ for a single-particle wavefunction ψ, introduce the boundary phases, or "twists", $\boldsymbol{\phi} = (\phi_x, \phi_y)$ via

$$\psi(\boldsymbol{x} + \boldsymbol{L}_x) = e^{i\phi_x}\psi(\boldsymbol{x}),\ \psi(\boldsymbol{x} + \boldsymbol{L}_y) = e^{i\phi_y}\psi(\boldsymbol{x}). \tag{3.10}$$

Then a unitary transformation of the form

$$\chi(\boldsymbol{x}) = e^{-i(\phi_x x/L_x + \phi_y y/L_y)}\psi(\boldsymbol{x}) \tag{3.11}$$

will transfer the twist angles to the Hamiltonian. Under the change of basis (3.11), Kane and Mele's model Hamiltonian [39] $H = H_0 + H_{SO} + H_R$ (see Section I) becomes (suppressing spinor indices)

$$H \to \mathcal{H}(\boldsymbol{\phi}) = \sum_{\langle ij\rangle} c_i^\dagger \left[t + i\lambda_R(\boldsymbol{s} \times \hat{\boldsymbol{d}}_{ij})_z\right] c_j e^{-i\boldsymbol{\Delta}\cdot\boldsymbol{d}_{ij}} + i\lambda_{SO}\sum_{\langle\langle ij\rangle\rangle} \nu_{ij} c_i^\dagger s^z c_j e^{-i\boldsymbol{\Delta}\cdot\boldsymbol{d}_{ij}} + \sum_i (\lambda_v \xi_i + w_i) c_i^\dagger c_i, \tag{3.12}$$

where $\boldsymbol{\Delta} = (\phi_x/L_x, \phi_y/L_y)$, $\boldsymbol{d}_{ij}$ is still the vector $i \to j$, and we have added a random term for on-site disorder, w_i, drawn from the Gaussian distribution of zero mean and standard deviation σ_w. It is now clear that under time reversal,

$$\Theta\mathcal{H}(-\boldsymbol{\phi})\Theta^{-1} = \mathcal{H}(\boldsymbol{\phi}), \tag{3.13}$$

since the extra phase factors in $\mathcal{H}$ change sign under complex conjugation.[2] Since this directly parallels equation (3.8), the $\mathbb{Z}_2$ invariant D of the Brillouin zone passes directly to twist space:

$$D_\phi = \frac{1}{2\pi}\left[\oint_{\partial(ETZ)} d\boldsymbol{\phi}\cdot\boldsymbol{\mathcal{A}} - \int_{ETZ} d^2\phi\, \mathcal{F}\right] \mod 2, \tag{3.14}$$

[2]Note that this means $\mathcal{H}$ is not generically time-reversal invariant. Indeed, there are only four boundary conditions that respect time-reversal, namely those for which ψ picks up a real phase upon translation around each cycle of the honeycomb lattice. These four correspond to the TRIM of the clean system in the calculation of D.

with ETZ for Effective Twist Zone, i.e., ETZ $= \{\boldsymbol{\phi} \,|\, 0 \le \phi_x \le \pi,\, -\pi < \phi_y \le \pi\}$. Note that there is an independent Chern parity D_ϕ for each Kramers-degenerate band pair separated from the rest of the spectrum by a gap at all $\boldsymbol{\phi}$: for such an isolated pair, $\boldsymbol{\mathcal{A}}$ and $\mathcal{F}$ are defined as in (3.7) and after, with $\boldsymbol{k} \to \boldsymbol{\phi}$ and $u \to \chi$.

In order to make contact with the band-structure definition, we note that if there is no disorder in the Hamiltonian (i.e., $\sigma_w = 0$), there are discrete translational symmetries within the $L_x \times L_y$ supercell that induce additional non-Kramers degeneracies at some points in twist space. With such degeneracies only the total Chern parity of all the degenerate states is well-defined. Disorder, as discussed in the following section, breaks all degeneracies resulting from translational invariance, leaving only separated band pairs, each of which has its own Chern parity. We now discuss to what extent Chern parities can be connected to observable quantities in a finite system.

3.2.3 Charge pumping cycles in time-reversal-invariant systems

The total Chern number in a finite IQHE system can be interpreted as measuring the number of charges pumped when the flux through one noncontractible circle on the torus increases adiabatically by one flux quantum. Briefly, one of the boundary phases corresponds to this driving flux, and the average over the other can be shown to yield the pumped charge [61]. The idea of "$\mathbb{Z}_2$ pumping" suggested by Fu and Kane [23] is the following: in a finite cylinder with boundaries, the operation of increasing the phase ϕ_x (in the periodic direction, around the cylinder) from 0 to π, corresponding to a magnetic flux of one-half flux quantum through the cylinder, has the following effect in the topological insulator. The values $\phi_x = 0$ and $\phi_x = \pi$ are special because, unlike general values, they are consistent with time-reversal invariance. At these special fluxes there are gapless states at the Fermi level that are localized near the edges because of the bulk gap. Fermi statistics requires that these states lie in Kramers doublets. If at zero flux, the Kramers doublet at one edge is partially occupied (has one state occupied), then the operation of changing the flux changes its occupancy to either double or zero occupancy. Since total charge is conserved, this requires a flow of $\mathbb{Z}_2$ from one boundary to the other: the final state differs from the initial state.

We now give an alternate definition of the topological insulator in terms of *cyclic* pumping of ordinary charge. This definition is mathematically equivalent to the definition of 3.2.2 based on treating the finite system as a supercell. In order to describe a closed pumping cycle, we need to add a second stage to the process of increasing boundary phase ϕ_x from 0 to π: although both these phases are consistent with time-reversal invariance, physical properties, like the occupancy of an edge doublet, are not identical at these different values of ϕ_x (even at the same ϕ_y). Hence the charge pumped in this process is not automatically well-defined. The only requirement on the second stage is essentially that it return the system to its original state *without applying a time-reversal-breaking flux*. While this definition precisely reproduces the supercell definition above, it should be noted that it is slightly different from pumping in the IQHE since, in some cases, the second stage requires changing the system's

Hamiltonian and not just the boundary phase.

Although the number of charges pumped is dependent not only on the first stage but on the second stage as well, whether this is an even or odd number is entirely determined by the first stage, as we now show. A closed pumping cycle is shown in Fig. 3.3. The original physical system's ETZ is the first stage of the cycle: the Hamiltonians are functions of ϕ_x from 0 to π and ϕ_y from $-\pi$ to π, with time-reversal constraints that act on the boundary circles at $\phi_x = 0$ and $\phi_x = \pi$. If ϕ_x takes one of these values, then the Hamiltonian at ϕ_y is time-reversal conjugate to that at $\phi'_y = -\phi_y$. The second stage can be any continuous change of the Hamiltonians that takes the $\phi_x = \pi$ system back to the $\phi_x = 0$ system and always satisfies the conjugacy condition between ϕ_y and ϕ'_y. This is the key difference between the second stage and the first stage: for intermediate values $0 < \phi_x < \pi$, there is no such conjugacy condition. The physical interpretation is that the second stage should be possible without introducing flux through the first noncontractible circle. The Fukui-Hatsugai algorithm [26] reviewed in the following section can be used to "choose a gauge" and determine a second stage with desired Chern numbers, which we have done as a check for graphene.

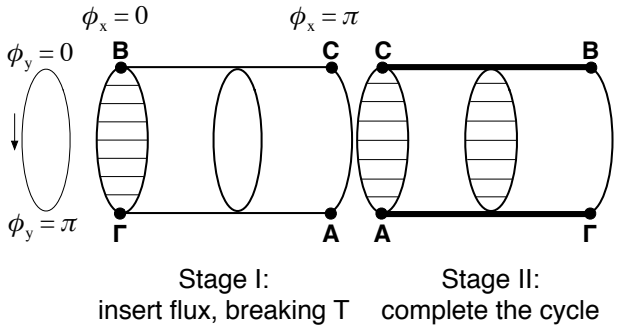

Figure 3.3: Graphical representation of charge pumping cycle for Chern parities. The first stage takes place on the ETZ (as in equation (3.14)), and the flux ϕ_x increases adiabatically from 0 to π. In the second stage the Hamiltonian at $(\phi_x = \pi, \phi_y)$ is adiabatically transported through the space of Hamiltonians to return to the Hamiltonian at $(\phi_x = 0, \phi_y)$. The difference between the second stage and the first is that at every step of the second stage, the Hamiltonians obey the time-reversal conditions required at $\phi_x = 0$ or $\phi_x = \pi$. The bold lines indicate paths along which all Hamiltonians are time-reversal invariant, and the disk with horizontal lines indicates, as before, how pairs of points in the second stage are related by time-reversal.

Now the torus shown in Fig. 3.3 has one Chern integer for each isolated band pair. Summing over occupied bands gives the amount of charge pumped in the cycle. Although this integer charge depends on the second stage, its parity is solely determined by the first stage, i.e., the physical system. In particular, for the ordinary insulator there is some closed cycle that pumps zero charge, while for the topological insulator there is some closed cycle

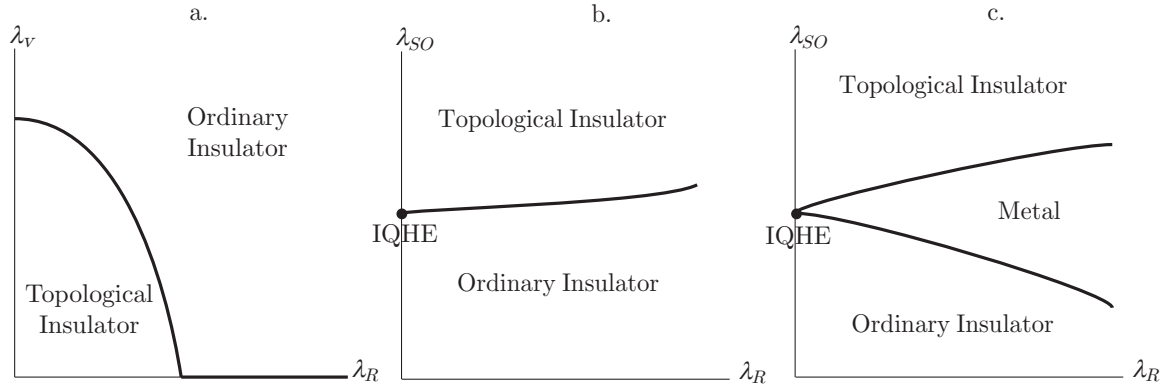

Figure 3.4: Schematic phase diagram. (a) Phase diagram for a clean system, with fixed t and $\lambda_{SO} \neq 0$ (after Kane and Mele [40]). (b) Again a clean system, now with fixed t and $\lambda_v \neq 0$. (c) The expected form of the phase diagram at nonzero disorder (we run all simulations at fixed λ_v). The phase boundary in (b) opens up into a metallic phase, closing only when $\lambda_R = 0$, where there should be an IQHE transition.

that pumps unit charge. These results follow from the same proof as for the band structure case in Ref. [56]: one shows that the differences in resulting Chern integers between any two second stages are even. The pumping definition gives a different physical picture for the "contractions" introduced there; instead of contracting the EBZ to a sphere, here the ETZ is contracted to a torus by adding the second stage. The technical reason that these two constructions are equivalent is that, since the appropriate spaces of Hamiltonians are contractible (i.e., have $\pi_1 = 0$), the two closed manifolds, the torus and the sphere, both have the same topological invariants, namely integer-valued Chern numbers.

The topological insulator in disordered systems has been studied previously by locating the transition between topological and ordinary insulators as a point or region where the localization length of single-electron eigenstates diverges. The existence of the topological insulator is inferred from the existence of this transition region (or alternately from the edge states in the topological insulator phase). In principle this approach could give a different result from ours, in that our definition probes the existence not just of extended states but specifically of extended states that contribute to the pumping of charge, or alternately that can give rise to edge states. The same distinction arises in the quantum Hall effect, where looking for extended states of nonzero Chern number is a more direct probe of quantum Hall physics than considering the inverse participation ratio, for example, which would detect extended states of zero Chern number in addition to topological states. However, as in the quantum Hall case, we find that the phase boundaries from our Chern-parity definition are consistent with those obtained from calculations of the localization length [64, 63].

3.3 Graphene Model and Numerics

3.3.1 The phase diagram of the disordered graphene model

It is useful to review some general expectations before applying the definitions of the previous section to study a disordered version of the graphene model. In two-dimensional systems with time-reversal invariance and no spin-orbit coupling, even very weak disorder will localize electron wavefunctions, so that these systems do not ever conduct in the thermodynamic limit. In the presence of a magnetic field, as in the IQHE, there are isolated energies with extended states, but no finite-width range of energies with extended states, and hence no true metallic phase. Hikami *et al.* showed [34] that disordered two-dimensional systems *with spin-orbit coupling* can nevertheless support a metallic phase, referred to as a "symplectic metal". They calculated ladder diagrams for the problem of random potential scattering in 2D in order to obtain a renormalization-group equation for the resistance, and found that, in the presence of strong spin-orbit scattering without magnetic scattering (so that the system is time-reversal invariant), the resistance flows to zero, indicating extended states and a metallic phase. In the presence of strong magnetic scattering (i.e., the IQHE), by contrast, the resistance does not flow at that order, and at next order grows (see also Ref. [33]).

In the case studied here, any metallic phase induced by disorder would presumably appear in a region around the parameter set that closes the (clean) gap, as depicted schematically in Fig. 3.4. At $\lambda_R = 0$ the z component of spin is a good quantum number (s^z commutes with H), so the system reduces to two copies of the Haldane model [29], which has a quantum Hall plateau transition with no metallic phase. Hence when $\lambda_R = 0$ there should be a direct transition between insulators even with disorder.

3.3.2 Lattice Implementation

For numerical work we use the algorithm of Fukui and Hatsugai, which we review here [26]. The formula (3.9) for D requires a gauge choice for the Hamiltonian eigenstates at each $\boldsymbol{\phi}$ on the two boundaries of the half-torus $0 \leq \phi_x \leq \pi$, $-\pi < \phi_y \leq \pi$. That is, the "field strength" $\mathcal{F}$ is gauge invariant, but the gauge potential $\boldsymbol{\mathcal{A}}$ is not. The eigenstates form Kramers pairs related by time reversal, and the gauge choice must respect this constraint.

Now, at the time-reversal invariant points $\boldsymbol{\phi} = (0,0)$, $(0,\pi)$, $(\pi,0)$, and (π,π), the solid points in Fig. 3.5, the spectrum is degenerate, with two states at each energy. The gauge condition requires that Θ interchange the two with a phase factor $e^{\pm i\pi/2}$ (so that $\Theta^2 = -1$ as required for single-fermion states). Numerical diagonalization will not, in general, return eigenvectors that obey this condition, but we can force them to do so as follows: choosing one of the two members of each Kramers pair at energy $\varepsilon_{2n-1} = \varepsilon_{2n}$ and calling that vector χ_{2n-1}, we discard the other and replace it by

$$\chi_{2n} = \Theta\chi_{2n-1}. \tag{3.15}$$

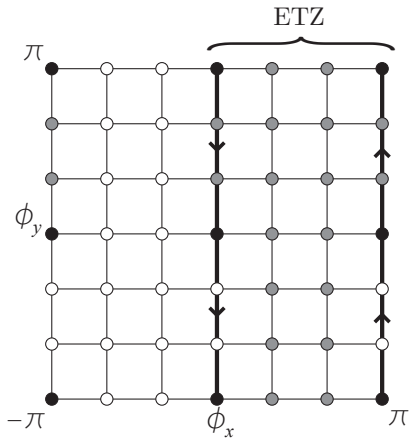

Figure 3.5: Twist Space. The bold lines indicate the boundaries of the "Effective Twist Zone", the region we integrate (or sum) over to calculate the Chern parity. The arrows indicate the direction to perform the sum over the boundary terms, and the lattice sites in gray indicate those for which the Hamiltonian eigenvectors are independently specified. That is, time reversal symmetry determines the eigenvectors on the white sites once those at the gray sites are found.

On the rest of the boundary, eigenvectors can be chosen freely on $0 < \phi_y < \pi$. On $-\pi < \phi_y < 0$ the algorithm takes

$$\chi_n(-\boldsymbol{\phi}) = \Theta\chi_n(\boldsymbol{\phi}). \tag{3.16}$$

In summary, the algorithm leaves alone the results of numerical diagonalization at all the gray points in Fig. 3.5, and by hand enforces the gauge condition on the rest of $\partial(\text{ETZ})$.[3]

With the eigenstates fixed at each point on the ETZ, we follow Fukui and Hatsugai and construct $U(1)$ parallel transporters on the links, as

$$U_x(\boldsymbol{\phi}) = \frac{g_x}{|g_x|},\ g_x = \det \boldsymbol{\chi}^\dagger(\boldsymbol{\phi})\boldsymbol{\chi}(\boldsymbol{\phi} + \hat{\boldsymbol{x}}) \tag{3.17}$$

and U_y similarly. Like $\boldsymbol{u}$ in (3.7), $\boldsymbol{\chi}$ is a matrix built from occupied state vectors, and $\hat{\boldsymbol{x}}$ translates by one link in the ϕ_x direction. In the continuum limit g should approach a pure phase, but for non-zero lattice spacing it will in general have $|g| \leq 1$, since the occupied subspace of interest will not embed in the total Hilbert space in the same way at every lattice point.

In the end, the only retained information will be the variation of the relative phases (hence the definition of U), which can be captured by choosing a lattice constant so small that the

[3]The points (ϕ_x, π) and $(\phi_x, -\pi)$ are not physically distinct; as pointed out earlier, a gauge transformation relates $\mathcal{H}(\phi_x, -\pi)$ to $\mathcal{H}(\phi_x, \pi)$ as given, for example. We therefore impose $\mathcal{H}(\phi_x, -\pi) \equiv \mathcal{H}(\phi_x, \pi)$ in the calculation. Alternatively, we could introduce boundary terms along $\phi_y = \pi, -\pi$ to compensate for the discrepancy.

phase field varies slowly over one link. However, the scale of variation will presumably differ for different disorder realizations, and we would like a way to diagnose this and throw out those realizations for which fast variation makes the calculation unreliable. The phase is periodic, and if it were to wind through 2π over the distance of one link the algorithm would miss this fact, so the relative phase will not provide a good diagnostic. As a proxy, we choose to cull out disorder realizations that result in small determinants in most simulations, since a small overlap between adjacent occupied eigenspaces indicates rapid variation. Of course, this filtering could introduce a selection bias into the results; these effects are within the statistical uncertainty of our analysis (in particular, within the error bars of Fig. 3.9).

Associated with the transporter on each link is a gauge potential $A_{x,y} = \log U_{x,y}$. This A is pure imaginary, and the logarithm is defined to return the branch $A/i \in (-\pi, \pi)$. Associated with the transport around each plaquette is a flux

$$F(\boldsymbol{\phi}) = \log U_x(\boldsymbol{\phi}) U_y(\boldsymbol{\phi} + \hat{\boldsymbol{x}}) U_x^{-1}(\boldsymbol{\phi} + \hat{\boldsymbol{y}}) U_y^{-1}(\boldsymbol{\phi}), \tag{3.18}$$

again satisfying $F/i \in (-\pi, \pi)$. With these definitions in hand, the lattice $\mathbb{Z}_2$ invariant corresponding to D_ϕ (3.14) is

$$D_L = \frac{1}{2\pi i} \left[\sum_{|\in \partial(ETZ)} A_y - \sum_{\square \in ETZ} F \right] \mod 2. \tag{3.19}$$

Of course, the sum over the boundary should have the same orientation as the corresponding contour integral in (3.14); in Fig. 3.5, this means the sum on the left boundary should carry a minus sign, following the arrows. As mentioned previously, this formalism has the desirable property of guaranteeing that $D_L = 0 \text{ or } 1$, but using too coarse a mesh can return the wrong value.

3.3.3 Numerical Results

As noted after equation (3.14), the number of nondegenerate Kramers pairs in a disordered system will generically be extensive, and we can use equation (3.19) to calculate the Chern parity of each pair separately. Fig. 3.6 shows the results of such calculations on a 4×6 lattice to present a picture of the three phases: normal insulator, symplectic metal, and topological insulator. In all phases, there are Kramers pairs with $D_\phi = 1$ in the lower half of the energy spectrum, as indicated by the bars. However, in the normal insulator (λ_{SO} well below the transition, or gap closure) there are an *even* number of such pairs in the occupied half of all realizations, so that the overall Chern parity is even. The presence of a very small number of realizations (2 of 215 for a 6×8 lattice) with odd parity indicates either the tail of the disorder-broadened transition or, more likely, that the weak filter applied for these simulations failed to catch all realizations with rapidly varying link variables.

When λ_{SO} is large, almost all realizations have an odd number of $\mathbb{Z}_2$-odd Kramers pairs (as with the small-λ_{SO} case, 2 realizations do not follow the rule), and in the region near the

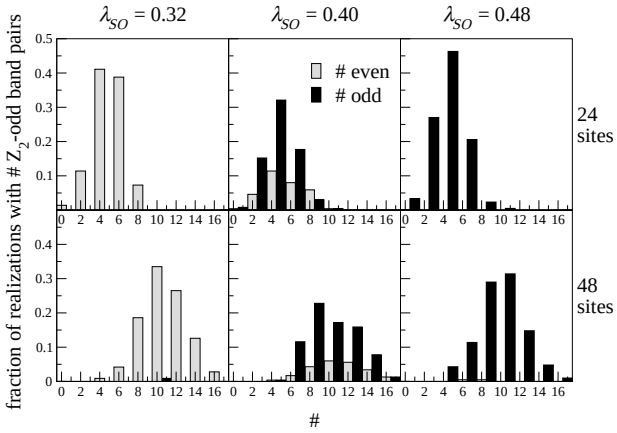

Figure 3.6: Distribution of Chern parities for band pairs. The bar heights represent the fraction of disorder realizations (out of ~ 200 trials) that have a given number # of band pairs with $D_L = 1$ in the occupied (half-filled) subspace. Those with an even number # will have an overall $D_L = 0$, those with an odd number will have overall $D_L = 1$. All these simulations were done with $t = -1$, $\lambda_R = \lambda_v = 1$, and $\sigma_w = 0.3$. Reading across, λ_{SO} increases and the system transitions from all realizations having even parity at small λ_{SO} to odd parity at large λ_{SO}. Reading down, doubling the system size doubles the total number of Kramers pairs and roughly doubles the number of $\mathbb{Z}_2$-odd pairs.

transition of the clean system there are instances of both types. The presence of disorder causes the "gap" to close at different values of λ_{SO} for different realizations, and also at different energies. The latter fact means that extended states are present throughout a finite spread of energies, while the former means that the metallic state exists over a finite region of parameter space.

For an integer quantum Hall system, Yang and Bhatt [90] have shown how to extract the localization length exponent ν from such calculations, in their case the Chern numbers of Landau sublevels. Specifically, sublevels with non-zero Chern number contain extended states, which should only occur at isolated energies in the IQHE. Therefore, the number N_c of such sublevels should decrease as the system size N_s increases, and in fact $\langle N_c \rangle \propto N_s^{1-1/2\nu}$, where $\langle\rangle$ indicates an average over disorder realizations. A similar approach for the QSH system here should reveal $\langle N_D \rangle \propto N_s$, where N_D is the number of $\mathbb{Z}_2$-odd Kramers pairs, since we expect a stable metallic band of energies in the thermodynamic limit (as observed by Obuse *et al.* [63] and Onoda *et al.* [64]). With enough data and large enough systems, finite-size-induced broadening of the edges of this band should also make ν accessible via a subleading term in the scaling. We find that larger system sizes require a finer mesh in twist space for these Kramers-pair-resolved simulations to return stable results, so that the requirements quickly outstrip our resources. Nevertheless, comparison of the two rows in Fig. 3.6 indicates that the location of the mean roughly doubles. This is what would be

expected for the middle case given the above considerations; the mean for the topological insulator (the right-hand panels in Fig. 3.6) should grow more slowly, as in the case studied by Yang and Bhatt.

The total phase of the system, given by the total Chern parity, is more relevant to possible measurements than the Chern parity of each Kramers pair. The former maintains its meaning if we consider the ground state wave function of the many-electron system, formed as a Slater determinant of the single-electron states we use here. There is also a computational benefit to calculating D_ϕ for the whole occupied subspace rather than for individual pairs, as well — at larger system sizes, and also at stronger disorder, the link variables for the half-filled subspace vary much more slowly than those for the individual Kramers pairs, making the calculation more robust. For these reasons the remaining plots in this paper depict only the Chern parity of the half-filled system.

To show that a metallic region of non-zero extent in parameter space exists in the thermodynamic limit, we need to verify that the mixed-phase region does not shrink to zero as we increase the system size. Figure 3.7(a) shows that as we make the system larger, the transition region certainly does not get narrower, and in fact the largest system size seems to have the broadest transition.

We can quantify the scaling of the metallic region's width with system size by assuming a simple one-parameter scaling form for the curves in Fig. 3.7(a) and defining the width of the curve to be proportional to the reciprocal of the maximum slope: width $\sim 1/\text{slope}$. With sufficient data, one could expand the approximately linear region near the middle of the transition in a power series with a few coefficients as fit parameters. Since our simulation data are limited, we opt instead to assume the form $(\tanh \alpha(s) + 1)/2$, which has roughly the right shape. If $\alpha = m(\lambda_{SO} - \lambda^*_{SO})$, then m is exactly the maximum slope we want. Figure 3.7(b) plots the data versus the best-fit scaling variable α for each system size; the points appear to fill out a smooth curve, justifying the scaling hypothesis. The best fit for m and λ^*_{SO} is determined by minimizing a weighted χ^2 statistic [91]. In particular, we assume that the results ($D_L = \pm 1$) of independent simulations at a fixed parameter set are distributed binomially, and that for each system size there are the same number of trials at each value of λ_{SO} (which is roughly true). In that case, the variance of the distribution can be estimated as $\sigma^2 \propto p(1-p)$ [15], where p is the fraction of disorder realizations returning $D_L = 1$, and then $1/\sigma^2$ is an appropriate weight for the statistical test. The error bars in Figs. 3.7 and 3.8 are also assigned based on a binomial model of the data [91].

By contrast, at $\lambda_R = 0$ the Hamiltonian (3.5) reduces to two copies of the Haldane model and so should exhibit the quantum Hall plateau transition, which looks like a step function at zero temperature in the thermodynamic limit. In Fig. 3.8(a) the width of the transition region shrinks as the system size grows, consistent with the prediction.

More quantitatively, Pruisken [66] has shown in a renormalization group framework that the functional form of the crossover for the IQHE looks like

$$p(L, B) = f(\alpha), \quad \alpha \propto L^{1/\nu}(B - B^*), \tag{3.20}$$

where L is the linear size of the finite sample, B is the applied magnetic field, and ν is

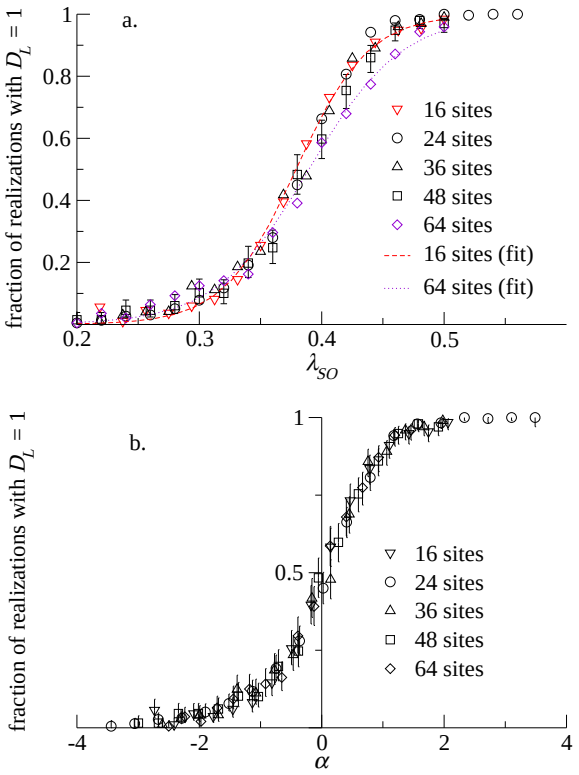

Figure 3.7: (a) At finite λ_R, the "metallic" region persists as the system size grows, and even broadens in the case shown here ($t = -1$, $\lambda_R = \lambda_v = \sigma_w = 1$). We identify the metallic region as those values of λ_{SO} for which some, but not all, disorder realizations have $D_L = 1$. As explained in the text, the fits to the simulation data have the form $(\tanh \alpha + 1)/2$, with $\alpha = m(\lambda_{SO} - \lambda^*_{SO})$. (b) The scaling collapse of the data in (a), based on the best fit m and λ^*_{SO}) for each system size. The error bars represent 95% confidence intervals assuming a binomial distribution of outcomes for each λ_{SO}.

again the localization length exponent. The function p could be either the longitudinal or transverse conductivity in the IQHE. Therefore, $p(\infty, B^* \pm \varepsilon) = f(\pm\infty)$, i.e., the transition is sharp in the thermodynamic limit. In our system the analogous parameter to B is λ_{SO}: in the Haldane model, the spin-orbit coupling breaks time-reversal invariance locally, like B does in the IQHE. The width of the transition region in B is governed by the way the Landau level energies respond to changing B, and the width of the transition region in our model (for fixed λ_v) is determined by the response of the gap to changing λ_{SO}. Since both responses are linear, we expect that the appropriate scaling variable will be $\alpha \propto L^{1/\nu}(\lambda_{SO} - \lambda^*_{SO})$.

This form would allow us to extract the exponent ν from the scaling of the maximum slope with system size for large systems (there are corrections at small system sizes). Again making

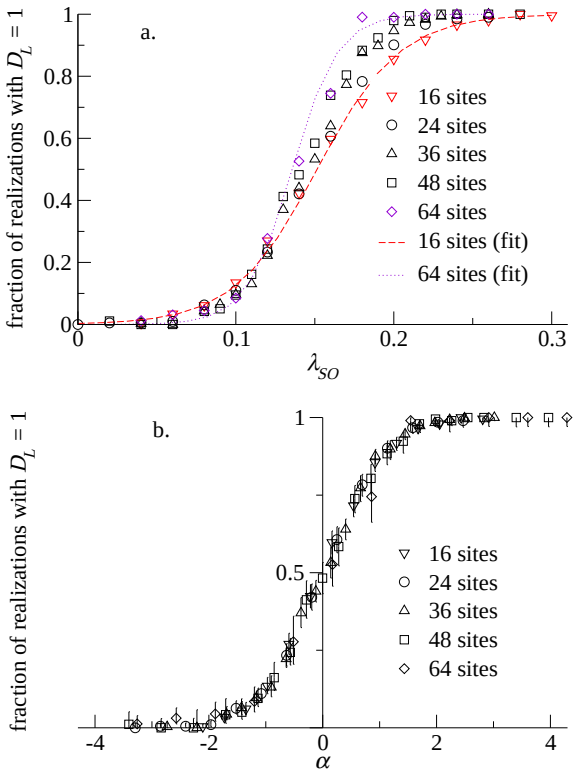

Figure 3.8: (a) At $\lambda_R = 0$, the metallic region gets narrower as the system size increases. As in Fig. 3.7, $t = -1$, $\lambda_v = \sigma_w = 1$. (b) The scaling variable α again comes from fitting to a tanh, and error bars represent 95% confidence intervals.

the fit to a tanh described above, Fig. 3.8(b) shows the scaling form, and Fig. 3.9 shows the scaling of width with linear system size. In particular, a regression gives $1/\nu = 0.78 \pm 0.03$, to be compared with the accepted value of $1/\nu \approx 0.42$. That is far from good agreement, but the observed scaling should not be taken as implying a new universality class. (For reference, the network-model work by Obuse *et al.* [63] found $1/\nu \approx 0.37$, and Onoda *et al.* [64] recently found a value $1/\nu \approx 0.63$) First, there should be finite-size corrections to the simple scaling assumed for the small systems considered here. Second, the scaling form is in principle different for different geometries, and the simulations were done for systems of varying aspect ratio (1 to 1.5). Nevertheless, it is clear that the qualitative behavior at $\lambda_R = 0$ is as expected, showing no metallic phase, and the behavior at $\lambda_R = 1$ is consistent with the presence of a metallic region.

Finally, by varying λ_R and noting the λ_{SO} values that mark the edges of the transition region for each λ_R, we can map out the phase diagram of the Hamiltonian (3.5) on a system of fixed size, as in Fig. 3.10. The widths obtained this way are in rough agreement with

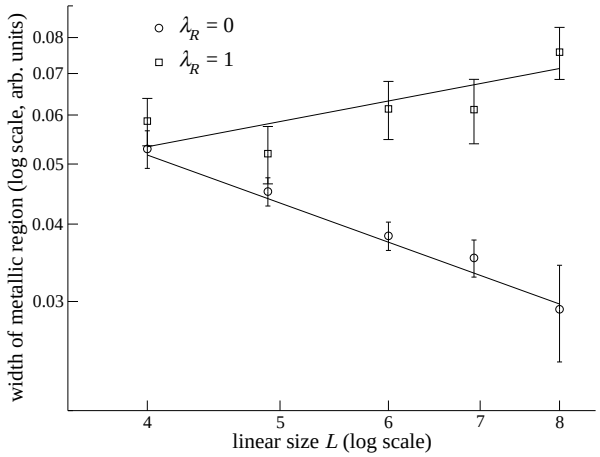

Figure 3.9: Width of the metallic transition region as a function of linear size L. The width is determined by the reciprocal of the maximum slope of the simulation data in Figures 3.7 and 3.8, and the linear size is taken to be the square root of the number of sites.

the scaling analysis outlined above, which returns the behavior of the width and not its normalization. As the finite-size scaling results in Figures 3.7 and 3.8 show, this phase diagram overestimates the width of the metallic phase at $\lambda_R = 0$, which is really zero. Together, these simulations confirm the expectation of Fig. 3.4 within the accuracy of our computational methods.

The same numerical methods could be used to study different forms of spin-orbit coupling quantitatively. The phase diagram of the model we have studied is reasonably simple in that, when the spin-independent part of the Hamiltonian is fixed and the spin-orbit couplings are not too strong, the intrinsic SO coupling pushes the system toward the topological insulator, and the Rashba coupling pushes the system toward the ordinary insulator. For more complicated forms of the spin-orbit coupling, the same analysis could be carried out to find the phase diagram, but it is not always possible to parametrize spin-orbit coupling with two parameters as for the graphene model. Also note that in an experiment, modifying the system using a perturbation such as a gate voltage is likely to modify both the spin-orbit and spin-independent parts of the Hamiltonian, so that a purely spin-orbit phase diagram would be insufficient. However, an important and potentially universal feature of the graphene phase diagram is that, *unless there is an extra conservation law* such as the S_z conservation when Rashba coupling is absent, a metallic region always appears between ordinary and topological insulators in the presence of disorder. The existence of this metallic region is an important difference between the spin Hall effect and the integer quantum Hall effect.

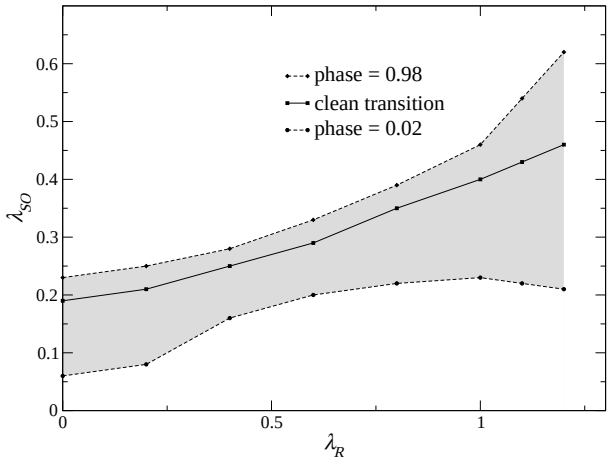

Figure 3.10: Approximate width of metallic region for a 4×6 lattice (fixed $\lambda_v = \sigma_w = 1$). The dashed curves indicate the parameter values at which 98% and 2% of the disorder realizations have $D_L = 1$. This underestimates the true width of the "metallic" region but hopefully avoids some amount of the inevitable error due to small system size. This diagram should offer a reasonable approximation to the thermodynamic (infinite-system) phase diagram away from $\lambda_R = 0$. At $\lambda_R = 0$, there should be no metallic phase, but a sharp transition between the two insulating phases in the thermodynamic limit. Given the plots in Fig. 3.7, it appears that finite size effects also reduce the width of the metallic phase at large λ_R.

3.4 Summary

Previous work [39, 40, 71, 23, 56] defined a $\mathbb{Z}_2$ topological invariant in infinite lattices that is similar to the TKNN invariant for the integer quantum Hall effect [81]. In disordered systems with boundaries, Fu and Kane [23] defined a topological invariant in terms of pumping of the occupancy of Kramers-degenerate edge states. We have given a definition of a topological invariant valid for disordered systems without boundary, i.e., without appeal to edge states. The "Chern parity" can be thought of as describing either a finite system with boundary phases or an arbitrarily large supercell in an infinite lattice system with well defined wavevector. A physical effect of Chern parity is that it determines whether the amount of charge pumped in a certain type of closed pumping cycle is even or odd. The "ordinary" and "topological" insulator phases can be distinguished by this invariant as long as many-body effects do not prevent description of the ground state as a Slater determinant. Chern parity in the spin quantum Hall effect is the natural generalization of Chern number in the integer quantum Hall effect.

In a disordered system, the only degeneracies expected to survive are Kramers degeneracies at time-reversal invariant values of the boundary phases. In this case, each pair of states

related by Kramers degeneracies can be assigned its own Chern parity, and the overall Chern parity of all occupied state pairs determines the observable phase. The lattice algorithm for $\mathbb{Z}_2$ topological invariants laid out by Fukui and Hatsugai [26, 25] allugows numerical identification of the topological insulator phase in disordered QSH systems. Implementing this algorithm for the specific graphene model Hamiltonian of Kane and Mele [40] with added on-site disorder, we observe the ordinary and topological insulator phases in simulations. While the number of "odd" pairs (state pairs with odd Chern parity) varies with the disorder realization, there is an even number of odd pairs in the ordinary insulator, and an odd number of odd pairs in the topological insulator.

We find that a metallic phase opens up between the two insulating phases for generic spin-orbit coupling. This agrees with the prediction of Hikami *et al.* [34] that spin-orbit coupling can protect a 2D metallic phase from disorder, and confirms the simulation results of Onoda *et al.* [64] and Obuse *et al.* [63] The methods in this paper could in principle be used to study the three-dimensional case and confirm the argument in Ref. [24] that only one of the four invariants of a band structure [56, 72, 22] is stable to disorder. While there is now strong evidence [63] that the phase transitions in the 2D QSHE are, except for special points, in the previously studied symplectic metal-insulator class, there is as yet no numerical study of the phase transitions in three-dimensional topological insulators.

Chapter 4

Magnetoelectric polarizability and axion electrodynamics in crystalline insulators

The previous chapter dealt with the two-dimensional topological insulator, and showed that its topological invariant can be generalized to the case of finite samples with disorder. It also verified that the combination of spin-orbit coupling and disorder lead to a metallic phase in two dimensions. There are also three-dimensional topological insulators, which have a topological invariant that can be built from the 2d invariant already discussed. This (shorter) chapter uses an alternative formulation, in which the topological invariant is a physical response function, the magnetoelectric polarizability.

Magnetoelectric couplings in solids have recently been the subject of intense experimental and theoretical investigations [77, 21, 32]. A quantity of central importance is the linear magnetoelectric polarizability α_{ij} defined via

$$\alpha_{ij} = \frac{\partial M_j}{\partial E_i}\Big|_{\mathbf{B}=0} = \frac{\partial P_i}{\partial B_j}\Big|_{\mathbf{E}=0} \tag{4.1}$$

where E and B are electric and magnetic fields, P and M are the polarization and magnetization, and the equality can be obtained from commuting derivatives of an appropriate free energy. In general the tensor α has nine independent components, and can be decomposed as

$$\alpha_{ij} = \tilde{\alpha}_{ij} + \frac{\theta e^2}{2\pi h}\,\delta_{ij} \tag{4.2}$$

where the first term is traceless and the second term, written here in terms of the dimensionless parameter θ, is the pseudoscalar part of the coupling. Here we focus on magnetoelectric coupling resulting from the orbital (frozen-lattice) magnetization and polarization, which we label the orbital magnetoelectic polarizability (OMP).

In field theory, the pseudoscalar OMP coupling is said to generate "axion electrodynam-

ics" [82], and corresponds to a Lagrangian of the form ($c = 1$)

$$\Delta\mathcal{L}_{EM} = \frac{\theta e^2}{2\pi h}\mathbf{E}\cdot\mathbf{B} = \frac{\theta e^2}{16\pi h}\epsilon^{\alpha\beta\gamma\delta}F_{\alpha\beta}F_{\gamma\delta}. \tag{4.3}$$

An essential feature of the axion theory is that, when the axion field $\theta(\boldsymbol{r}, t)$ is constant, it plays no role in electrodynamics; this follows because θ couples to a total derivative, $\epsilon^{\alpha\beta\gamma\delta}F_{\alpha\beta}F_{\gamma\delta} = 2\epsilon^{\alpha\beta\gamma\delta}\partial_\alpha(A_\beta F_{\gamma\delta})$, and so 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 $\theta(\boldsymbol{r})$ appear.

A second essential feature is that electrodynamics is invariant under $\theta \to \theta + 2\pi$ [82]. In order to reconcile this peculiar fact with the phenomenology of the magnetoelectric effect, observe that the axion coupling can alternatively be described in terms of a *surface Hall conductivity* $\sigma_{\rm H}$ whose value $\theta e^2/2\pi h$ is determined by bulk properties, but only modulo the quantum e^2/h. More generally, at an interface between two samples, $\sigma_{\rm H} = (\theta_1 - \theta_2 + 2\pi r)e^2/2\pi h$, where the integer r depends on the details of the interface. Recall that, in general, a 2D gapped crystal has an integer TKNN invariant C in terms of which the its Hall conductivity is $\sigma_{\rm H} = Ce^2/h$ [81]. The "modulo e^2/h", or integer r, discussed above corresponds to modifying the surface or interface by adsorbing a surface layer of nonzero C.

When time-reversal (T) invariance is present, the TKNN invariants vanish, but other invariants arise that have been the focus of much recent work. In 2D there is a $\mathbb{Z}_2$ invariant [40] distinguishing "ordinary" from "$\mathbb{Z}_2$-odd" insulators, with "quantum spin Hall" states [39, 8] providing examples of the latter. In 3D there is a similar invariant [56, 72, 22] that can be computed either from the 2D invariant on certain planes [56] or from an index involving the eight T-invariant momenta [22]. If this is odd, the material is a "strong topological insulator" (STI). In the context of the OMP, note that T maps $\theta \to -\theta$; the ambiguity of θ modulo 2π then implies that T invariance is consistent with either $\theta = 0$ or $\theta = \pi$, with the latter corresponding to the STI [67]. Note that if T-invariance extends to the surfaces, these become metallic by virtue of topologically protected edge states, as observed experimentally for the $Bi_{0.9}Sb_{0.1}$ system [37]. If the surface is gapped by a T-breaking perturbation, then $\sigma_{\rm H} = e^2/h$ modulo $2e^2/h$ at the surface of a STI [82, 24, 67].

In the noninteracting case, a Berry-phase expression for θ has been given in terms of the bulk bandstructure by Qi, Hughes, and Zhang [67] by integrating out electrons in one higher dimension. Defining the Berry connection $\mathcal{A}_j^{\mu\nu} = i\langle u_\mu|\partial_j|u_\nu\rangle$ where $|u_\nu\rangle$ is the cell-periodic Bloch function of occupied band ν and $\partial_j = \partial/\partial k_j$, they obtain

$$\theta = -\frac{1}{4\pi}\int_{\rm BZ} d^3k\ \epsilon_{ijk}\,{\rm Tr}[\mathcal{A}_i\partial_j\mathcal{A}_k - i\frac{2}{3}\mathcal{A}_i\mathcal{A}_j\mathcal{A}_k] \tag{4.4}$$

where the trace is over occupied bands.

In the present letter, we first provide an alternate derivation of Eq. (4.4) for the OMP. Our derivation clarifies that θ is a polarizability and in fact describes the *entire* orbital contribution to magnetoelectric polarizability. That is, the orbital $\tilde{\alpha}$ vanishes in Eq. (4.2), and

the orbital contribution is a perfect pseudoscalar. The derivation follows from an extension [87] of the Berry-phase theory of polarization [43] to the case of slow spatial variations of the Hamiltonian. (Indeed, the OMP angle θ is a bulk property in exactly the same sense as electric polarization [65, 43].) We find that the OMP can be generalized to the interacting case and calculated from the many-particle wavefunction, even though Eq. (4.4) is not valid; this reflects a subtle difference between OMP and polarization. Explicit numerical calculations on model crystals are presented to validate the theory, establish the equivalence of Eq. (4.4) to the prior definition, and illustrate how a non-zero θ corresponds to a "fractional" quantum Hall effect at the surface of a magnetoelectric or topological insulator [82, 24, 67].

From Eq. (4.1) it is evident that the OMP can be viewed in several ways. (i) It describes the *electric* polarization arising from the application of a small *magnetic* field. (ii) It describes the orbital *magnetization* arising from the application of a small *electric* field. (iii) It also gives the (dissipationless) *surface Hall conductivity* $\sigma_{\rm H}$ at the surface of the crystal, provided that the surface is insulating. Note that (iii) follows from (ii): for a surface with unit normal $\hat{\mathbf{n}}$ and electric field $\mathbf{E}$, the resulting surface current $\mathbf{K} = \mathbf{M} \times \hat{\mathbf{n}}$ is proportional to $\mathbf{E} \times \hat{\mathbf{n}}$. There is an elegant analogy here to the case of electric polarization, where the surface charge of an insulating surface is determined, modulo the quantum e/S, by the bulk bandstructure alone (S is the surface cell area).

The above discussion suggests two approaches to deriving a bulk formula for the OMP θ. One is to follow (ii) and compute the orbital magnetization [78, 88] in an applied electrical field. We focus here on (i) instead, working via dP/dB. The modern theory of polarization starts from the polarization current $j_P = dP/dt$ under slow deformation of the Bloch Hamiltonian, and contains, to first order in d/dt, one power of the Berry curvature defined below [43]. Using semiclassical wavepacket dynamics, Xiao *et al.* [87] have shown how to compute the polarization current to second order and to incorporate slow spatial variations in the electronic Hamiltonian. For the case of an orthorhombic 3D crystal with M occupied bands in which the slow spatial variation occurs along the y direction in a supercell of length l_y, they obtain

$$\langle \Delta P_x^{(\rm in)} \rangle = \frac{e}{8} \int_0^1 d\lambda \int_{\rm BZ} \frac{d^3k}{(2\pi)^3} \int_0^{l_y} \frac{dy}{l_y} \, \epsilon_{ijkl} {\rm Tr}[\mathcal{F}_{ij}\mathcal{F}_{kl}] \tag{4.5}$$

for the change in the supercell-averaged inhomogeneously induced polarization that occurs as a global parameter λ evolves adiabatically from 0 to 1. Here indices $ijkl$ run over (k_x, k_y, y, λ), $\mathcal{F}_{ij} = \partial_i \mathcal{A}_j - \partial_j \mathcal{A}_i - i[\mathcal{A}_i, \mathcal{A}_j]$ is the Berry curvature tensor ($\mathcal{A}_\lambda^{\mu\nu} = i\langle u_\mu | \partial_\lambda | u_\nu \rangle$), and the trace and commutator refer to band indices.

Because $\mathcal{F}$ is gauge-covariant, the integrand in Eq. (4.5) is explicitly gauge-invariant; it is the non-Abelian second Chern class [60], so that Eq. (4.5) is path-invariant modulo a quantum $e/a_z l_y$, where a_z is the lattice constant in the z direction. Moreover, the λ integral can be performed to obtain an expression in terms of the non-Abelian Chern-Simons 3-form [60]. Thus,

$$\langle P_x^{(\rm in)} \rangle = -\frac{e}{2} \int_{\rm BZ} \frac{d^3k}{(2\pi)^3} \int_0^{l_y} \frac{dy}{l_y} \, \epsilon_{ijk} {\rm Tr}[\mathcal{A}_i \partial_j \mathcal{A}_k - \frac{2i}{3} \mathcal{A}_i \mathcal{A}_j \mathcal{A}_k] \tag{4.6}$$

where ijk now run only over (k_x, k_y, y). Here the integrand is not gauge-invariant, but the integral is gauge-invariant modulo the quantum $e/a_z l_y$.

We apply this result to study the polarization

$$\langle P_x^{(\mathrm{in})} \rangle = -\frac{Be^2}{2\hbar} \int_{\mathrm{BZ}} \frac{d^3k}{(2\pi)^3} \epsilon_{ijk} \mathrm{Tr}[\mathcal{A}_i \partial_j \mathcal{A}_k - i\frac{2}{3} \mathcal{A}_i \mathcal{A}_j \mathcal{A}_k] \tag{4.7}$$

induced by a magnetic field described by the inhomogeneous vector potential $\mathbf{A} = By\hat{\mathbf{z}}$ with $B = h/ea_z l_y$, i.e., a B-field along $\hat{\mathbf{x}}$ with one flux quantum threading the supercell. This has the effect of taking $k_z \to k_z + eBy/\hbar$, and this is the only y-dependence in the Hamiltonian, so that $|\partial_y u\rangle = (Be/\hbar)|\partial_{k_z} u\rangle$ and where ijk now run over (k_x, k_y, k_z). Using Eqs. (4.1,4.2) we arrive directly at Eq. (4.4).

There is an important geometrical relationship in this (noninteracting) derivation that applies equally well to the many-body case and gives a bulk interpretation of the 2π ambiguity in θ, whose surface interpretation was in terms of allowed surface IQHE layers. Polarization in a crystal is defined modulo the "quantum of polarization" [43] which, for the flux-threaded supercell of Eq. (4.7), is $\Delta P_x = e/a_z l_y$. Since the magnetic field is $B_x = h/ea_z l_y$, it follows that $\Delta(P_x/B_x) = e^2/h$. Hence the unit-cell-independent ambiguity of dP/dB results from the relationship in a finite periodic system between the unit-cell-dependent polarization quantum and the quantization of applied flux, and this relationship remains valid in the many-body case.

Before studying the OMP in a specific model, we discuss its symmetry properties and how to obtain it when Bloch states are unavailable, as in the many-particle case. Clearly the combination $\mathbf{E} \cdot \mathbf{B}$ in Eq. (4.3) is odd under T and under inversion P (although it is even under the combination PT). It is also odd under any improper rotation, such as a simple mirror reflection. This implies that $\theta = -\theta$ if the crystal has *any* of the above symmetries. This would force an aperiodic coupling to vanish, but since θ is only well-defined modulo 2π, it actually only forces $\theta = 0$ or π. Thus, one can obtain an insulator with quantized $\theta = \pi$ not only for T-invariant systems (regardless of whether they obey inversion symmetry), but also for inversion- and mirror-symmetric crystals regardless of T symmetry [24]. When none of these symmetries are present, one generically has a non-zero (and non-π) value of θ, but still retaining the simple scalar form of Eq. (4.3).

In an interacting system, the OMP should be obtained from the many-particle wavefunction. However, modifying Eq. (4.4) to the Abelian Chern-Simons integral over the many-body wavefunction fails [1], in important contrast to the case of the polarization (the integral of $\mathcal{A}$), where such a generalization works [65]. Instead the OMP can be found using the change in the many-body polarization due to an applied magnetic field to compute dP/dB, i.e., the many-body version of the supercell dP/dB calculation. This fact is important beyond computing θ with interactions, as it defines the topological insulator phase in the many-body

[1] Mathematically, obtaining a nontrivial second Chern or Chern-Simons integral depends on having a degenerate set of bands somewhere in parameter space, which is not the case for a gapped many-body wavefunction.

case more simply than before [48]. Like the IQHE, the topological insulator is defined via a response function (dP/dB) to a perturbation that, in the limit of a large system with periodic boundary conditions, is locally weak and hence does not close the insulating gap. In the IQHE, this response function is to a boundary phase (i.e., a flux that does not pass through the 2D system), while for the topological insulator, the defining response is to a magnetic flux through the 3D system.

In the remainder of this Letter, we demonstrate the above theory via numerical calculations on a tight-binding Hamiltonian that generates non-zero values of θ, then discuss experimental measurements of θ. We start with the model of Fu, Kane, and Mele [22] for a 3D topological insulator on the diamond lattice,

$$H_{FKM} = \sum_{\langle ij \rangle} t_{ij} c_i^\dagger c_j + i \frac{4\lambda_{SO}}{a^2} \sum_{\langle\langle ij \rangle\rangle} c_i^\dagger \boldsymbol{\sigma} \cdot (\boldsymbol{d}_{ij}^1 \times \boldsymbol{d}_{ij}^2) c_j. \tag{4.8}$$

In the first term, the nearest-neighbor hopping amplitude depends on the bond direction; we take $t_{ij} = 3t + \delta$ for direction [111] (in the conventional fcc unit cell of linear size a) and $t_{ij} = t$ for the other three bonds. The second term describes spin-dependent hopping between pairs of second neighbors $\langle\langle ij \rangle\rangle$, where $\boldsymbol{d}_{ij}^1$ and $\boldsymbol{d}_{ij}^2$ are the connecting first-neighbor legs and $\boldsymbol{\sigma}$ are the Pauli spin matrices. With $|\delta| < 2t$ and λ_{SO} sufficiently large, this model has a direct band gap of $2|\delta|$.

To break T we add a staggered Zeeman field with opposite signs on the two fcc sublattices A and B, $\boldsymbol{h} \cdot \left(\sum_{i \in A} c_i^\dagger \boldsymbol{\sigma} c_i - \sum_{i \in B} c_i^\dagger \boldsymbol{\sigma} c_i \right)$. We take $|\boldsymbol{h}| = m \sin\beta$ and choose $\boldsymbol{h}$ in the [111] direction; setting $\delta = m \cos\beta$ and varying the single parameter β keeps the gap constant and interpolates smoothly between the ordinary ($\beta = 0$) and the topological ($\beta = \pi$) insulator.

We have calculated the OMP angle θ using four different methods with excellent agreement (Fig. 4.1). First, we obtain θ from Eq. (4.4); this requires a smooth gauge for $\mathcal{A}$, which can be found using now-standard Wannier-based methods [52]. Results are shown for $\beta = \pi/4$ and $\beta = \pi/2$ (filled squares).

Next, we have calculated the polarization [43]

$$P_i = e \int_{BZ} \frac{d^3k}{(2\pi)^3} \mathrm{Tr}\, \mathcal{A}_i \,. \tag{4.9}$$

resulting from a single magnetic flux quantum in a large supercell. Varying the supercell size (and thereby B) allows us to approximate dP/dB, yielding the open squares in Fig. 4.1. The points in Fig. 4.1 are from the surface Hall response in a slab geometry, described below. Finally, to obtain the curve in Fig. 4.1, we also computed $\theta(\beta)$ from the second Chern expression [87, 67]

$$\theta = \frac{1}{16\pi} \int_0^\beta d\beta' \int d^3k\, \epsilon_{ijkl} \mathrm{Tr}[\mathcal{F}_{ij}(\mathbf{k}, \beta') \mathcal{F}_{kl}(\mathbf{k}, \beta')] \tag{4.10}$$

(derived above as Eq. (4.5)). Clearly, the various approaches are numerically equivalent.

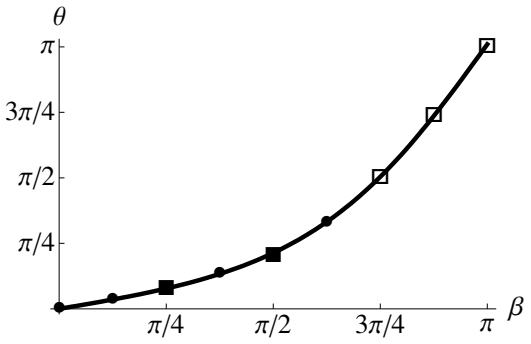

Figure 4.1: The magnetoelectric polarizability θ (in units of $e^2/2\pi h$). The filled squares are computed by the Chern-Simons form, Eq. (4.4). The open squares are dP/dB from Eq. (4.9). The points are obtained by layer-resolved $\sigma_{\rm H}$ calculations using Eq. (4.12). The curve is obtained from Eq. (4.10).

We now discuss the surface Hall conductivity, whose fractional part in units of e^2/h is just $\theta/2\pi$ [82]. Consider a material with coupling θ in a slab geometry that is finite in the $\hat{\mathbf{z}}$ direction and surrounded by $\theta = 0$ vacuum. The simplest interfaces will then lead to $\sigma_H = \theta e^2/(2\pi h)$ at the top surface and $-\theta e^2/(2\pi h)$ at the bottom surface, for a total σ_{xy} of zero. More generally, arbitrary surface quantum Hall layers change the total integer quantum Hall state, but not the fractional parts at each surface.

The spatial contributions to the Hall conductance in the slab geometry can be resolved as follows. The unit cell is a supercell containing some number N of original unit cells in the $\hat{\mathbf{z}}$ direction, with translational invariance remaining in the $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ directions. The TKNN integer for the entire slab is [81, 4]

$$C = \frac{i}{2\pi} \int d^2k \, \mathrm{Tr} \left[\mathcal{P} \epsilon_{ij} \partial_i \mathcal{P} \partial_j \mathcal{P} \right]. \tag{4.11}$$

Here i and j take the values k_x and k_y and $\mathcal{P} = \sum_\nu |u_\nu\rangle\langle u_\nu|$ is the projection operator onto the occupied subspace (ν runs over occupied bands). To find how different $\hat{\mathbf{z}}$ layers contribute to C, define a projection $\tilde{\mathcal{P}}_n$ onto layer n within the supercell, and compute

$$C(n) = \frac{i}{2\pi} \int d^2k \, \mathrm{Tr} \left[\mathcal{P} \epsilon_{ij} (\partial_i \mathcal{P}) \tilde{\mathcal{P}}_n (\partial_j \mathcal{P}) \right]. \tag{4.12}$$

The results, presented in Fig. 4.2, confirm that the surface layers have half-integer Hall conductance when $\beta = \pi$ in (4.8) and that the sign on each surface is switched by local T-breaking perturbations (in this example, a uniform Zeeman coupling in the surface layer).

To gain some insight into the microscopic origin of θ in the noninteracting case, using Eq. (4.4) we have calculated θ for a Hamiltonian that breaks PT (as well as P and T) by adding a weak, uniform (*i.e.*, not staggered) Zeeman coupling. For some values of β this lifts all degeneracies, enabling us to isolate the single-band and interband contributions to θ.

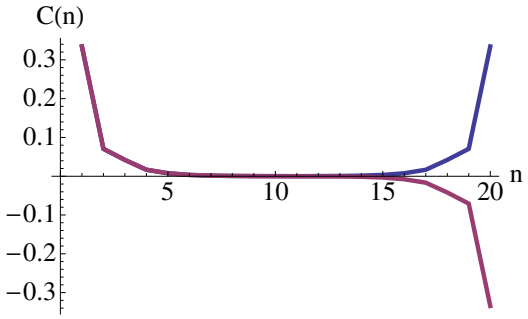

Figure 4.2: The layer-resolved Hall conductivity (in units of e^2/h) at $\beta = \pi$ in a slab of twenty layers, with $m = t/2$ and $\lambda_{SO} = t/4$, terminated in $(\bar{1}11)$ planes.

A single filled band can have nonzero θ only if there are more than two bands in total [57]. Because interband contributions are nonzero in general, θ is a property of the whole occupied spectrum, unlike polarization, which is a sum of individual band contributions.

Experimental detection of θ is more difficult for a topological insulator than for a generic magnetoelectric insulator because some T-breaking perturbation is needed to gap the surface state. Furthermore, a large surface density of states, as in BiSb, may complicate the measurement: while even a weak magnetic field will in principle lead to a gap and half-integer quantum Hall effect at each surface, the large number of filled surface Landau levels may make it difficult to isolate the half-integer part of surface σ_{H} in, e.g., optical conductivity. In the presence of broken discrete symmetries, as in antiferromagnets or multiferroics, the surface gap exists naturally and experiments are easier. For example, the theoretical methods of this paper could be used to compute the orbital part of the recently measured θ in Cr_2O_3 [32].

Interlude: magnetoelectric response of a finite material

The previous chapter gave a semiclassical derivation of the Chern-Simons term in the magnetoelectric polarizability in a bulk insulator. Before proceeding to the full, and fully quantum mechanical, derivation of the orbital magnetoelectric response in the next chapter, it is worth recording the result in a finite (insulating) sample of material. These results can also be found in Malashevich *et al.* [51]. To linear order in the magnetic field, the (explicitly Hermitian) Hamiltonian is

$$H_{tot} = H - \frac{1}{2}(\mathbf{j}\cdot\mathbf{A} + \mathbf{A}\cdot\mathbf{j}) = H - \frac{1}{2}\{j^i, A_i\}. \tag{4.13}$$

We use curly braces $\{,\}$ for the anticommutator. A convenient gauge for the vector potential $\mathbf{A}$ is

$$A_i = \frac{1}{2}\epsilon_{ijk}B^j r^k, \tag{4.14}$$

where ϵ_{ijk} is the fully antisymmetric Levi-Civita symbol. This gives

$$\partial_{B^j} H_{tot} = \frac{1}{4}\epsilon_{jab}\{j^a, r^b\} \tag{4.15}$$

The response of the electronic polarization will then be

$$\begin{aligned}\alpha^i_j = \partial_{B^j}\frac{\langle er^i\rangle}{\Omega} &= \frac{1}{\Omega}\sum_{n\,\mathrm{occ}}\langle n|er^i|\partial_{B^j}n\rangle + \text{c.c.} \\ &= \frac{e\epsilon_{jab}}{2\Omega}\,\mathrm{Re}\sum_{n\,\mathrm{occ}}\langle n|r^i G(E_n)\{j^a, r^b\}|n\rangle \\ &= \frac{e\epsilon_{jab}}{2\Omega}\,\mathrm{Re}\sum_{n\,\mathrm{occ},l}\frac{\langle n|r^i|l\rangle\langle l|\{j^a, r^b\}|n\rangle}{E_n - E_l + i\epsilon},\end{aligned} \tag{4.16}$$

where $G(E_n) = (E_n - H + i\epsilon)^{-1}$ is the Schrödinger Green function for the unperturbed Hamiltonian. We here take e to be the sign of the charge carrier, so it is negative for electrons; Ω is the volume of the material in question. Observe that the real part of the sum over occupied states l vanishes automatically, which is to say, there is effectively a factor

$1 - f(E_l)$, where $f(E)$ is the Fermi-Dirac distribution at zero temperature. Leaving the sum in this form is not helpful numerically, but it makes plain that α is a band-by-band sum of a single-band response function, a property we shall call band additivity.

In an infinite system, the position is unbounded, and in a system with periodic (Born-von Karman) boundary conditions it is either multivalued or discontinuous, implying that the expression we have used for the electronic polarization, $\langle e\mathbf{r}\rangle$, is of questionable validity in such cases. On the other hand, the current is perfectly well behaved, making it preferable to work with the current operator when possible. The two are related in the unperturbed system by

$$\mathbf{j} = \frac{e}{i\hbar}[\mathbf{r}, H], \tag{4.17}$$

or in the basis of energy eigenstates,

$$\langle l'|\mathbf{r}|l\rangle = \frac{i\hbar}{e}\frac{\langle l'|\mathbf{j}|l\rangle}{E_l - E_{l'}}. \tag{4.18}$$

Matrix elements between occupied and unoccupied states are finite, then, because of the gap, while potential problems only appear in those between the (possibly degenerate) occupied states, making it worthwhile to separate out those contributions to α that have such problematic matrix elements. Using Eq. (4.18) and $[r^i, r^j] = 0$ it follows that

$$\begin{aligned}\alpha^i_j =& \frac{e\epsilon_{jab}}{\Omega}\,\mathrm{Re}\sum_{\substack{n\,\mathrm{occ}\\ m\,\mathrm{unocc}}}\langle n|r^i|m\rangle\frac{\langle m|j^a(1-\rho)r^b|n\rangle + \langle m|r^b\rho j^a|n\rangle}{E_n - E_m}\\ &+ \frac{ie^2}{3\hbar\Omega}\delta^i_j\epsilon_{abc}\sum_{\substack{n,n',\\ n''\mathrm{occ}}}\langle n|r^a|n'\rangle\langle n'|r^b|n''\rangle\langle n''|r^c|n\rangle.\end{aligned} \tag{4.19}$$

Here ρ is the projector onto the occupied states, otherwise known as the Fermi-Dirac density matrix at zero temperature, and $1 - \rho$ is the complementary projector onto the unoccupied states. Note that the first sum looks very like Eq. (4.16) but with a factor of 2 difference. Remarkably, the troublesome (diagonal) matrix elements of $\mathbf{r}$ only contribute to the trace part of the magnetoelectric tensor, that is, to θ.

That is to say, the first term presented has a simple generalization to infinite and/or periodic systems. The second term does not, and is therefore the piece of most interest here. Let us rewrite it in terms of the density matrix as

$$(\alpha_{CS})^i_j = \frac{ie^2}{3\hbar\Omega}\delta^i_j\epsilon_{abc}\mathrm{Tr}\rho r^a\rho r^b\rho r^c = \frac{ie^2}{3\hbar\Omega}\delta^i_j\epsilon_{abc}\mathrm{Tr}\rho[\rho, r^a][\rho, r^b]\rho r^c \tag{4.20}$$

Ignoring the prefactors, this looks very much like a measure of volume, $\mathrm{Tr}(\rho\mathbf{r})\cdot(\rho\mathbf{r})\times(\rho\mathbf{r})$, and evidently gives a measure of the "quantum volume" of the state ρ. Unfortunately, it does not seem to be possible to express it solely in terms of the noncommutative position operator $\mathscr{r} = i[\rho, \mathbf{r}]$.

An alternative interpretation is as a polarization of Hall conductivities, since these are given by $\epsilon_{abc}\rho[\rho, r^a][\rho, r^b]\rho$. This formulation is essentially equivalent to reading off the response from the surface conductivity as was done in the last chapter. In the chapter to follow, the derivation of the orbital magnetoelectric response is given in full for an infinite (boundaryless) system. The results look very similar to those derived here, and allow us to conclude that the "quantum volume" introduced here corresponds precisely to the Chern-Simons integral derived earlier from the semiclassical equations of motion.

Chapter 5

General orbital magnetoelectric coupling in band insulators

5.1 Introduction

Understanding the response of a solid to applied magnetic or electric fields is of both fundamental and applied interest. Two standard examples are that metals can be distinguished from insulators by their screening of an applied electric field, and superconductors from metals by their exclusion of magnetic field (the Meissner effect). Magnetoelectric response in insulators has been studied for many years and is currently undergoing a renaissance driven by the availability of new materials. The linear response of this type is the magnetoelectric polarizability: in "multiferroic" materials that break parity and time-reversal symmetries, an applied electric field creates a magnetic dipole moment and a magnetic field creates an electric dipole moment, and several applications have been proposed for such responses. Such responses are observed in a variety of materials and from a variety of mechanisms [77, 21]. From a theoretical point of view, the most intriguing part of the polarizability is that due to the orbital motion of electrons, because the orbital motion couples to the vector potential rather than the more tangible magnetic field.

The orbital magnetoelectric polarizability has also been studied recently in non-magnetic materials known as "topological insulators." These insulators have Bloch wavefunctions with unusual topological properties, that lead to a magnetoelectric response described by an $\mathbf{E}\cdot\mathbf{B}$ term in their effective electromagnetic Lagrangians, [67] with a quantized coefficient. Qi, Hughes, and Zhang [67] (QHZ) gave a formula for the coefficient of this term. For the specific case of topological band insulators, their result reproduces earlier formulas for the relevant topological invariant, [22, 56, 72] but it is more generally valid: it describes a contribution to the magnetoelectric polarizability non just in topological insulators but in any band insulator. Their formula has a periodicity or ambiguity by e^2/h that is related to the possibility of surface quantum Hall layers on a three-dimensional sample and generalizes the ambiguity of ordinary polarization.

The same $\mathbf{E}\cdot\mathbf{B}$ coupling, known as "axion electrodynamics" and originally studied in the 1980s, [82] was obtained in a previous paper by three of the present authors [20] using a semiclassical approach [87] to compute dP/dB, the polarization response to an applied magnetic field. However, in a general material, that semiclassical approach leads to an explicit formula for only part of the orbital magnetoelectric polarizability, the part found by QHZ. [67] The remainder, which is generically nonvanishing in materials that break inversion and time-reversal symmetries, is expressed only implicitly in terms of the modification of the Bloch wavefunctions by the magnetic field.

In this paper, we develop a more microscopic approach that enables us to compute all terms in the orbital response explicitly in terms of the unperturbed wavefunctions, thereby opening the door to realistic calculations using modern band-structure methods (e.g., in the context of density-functional theory). Moreover, beyond its importance for computation, this expression clarifies the physical origins of the orbital magnetoelectric polarizability and resolves some issues that arose in previous efforts to describe the "toroidal moment" in periodic systems.

In the remainder of this introduction, we review some macroscopic features of the magnetoelectric response, while subsequent sections will be devoted mainly to a detailed treatment of microscopic features. The magnetoelectric tensor can be decomposed into trace and traceless parts as

$$\frac{\partial P^i}{\partial B^j}=\frac{\partial M_j}{\partial E_i}=\alpha_j^i=\tilde{\alpha}_j^i+\alpha_\theta\delta_j^i, \tag{5.1}$$

where $\tilde{\alpha}$ is traceless and

$$\alpha_\theta=\frac{\theta}{2\pi}\frac{e^2}{h} \tag{5.2}$$

is the trace part expressed in terms of the dimensionless parameter θ; α has the physical dimension of conductance. The trace is the most difficult term to determine, as its physical effects are elusive. It should be noted that equality between $\partial P^i/\partial B^j$ and $\partial M_j/\partial E_i$ only holds in the absence of dissipation and dispersion, which describes the low frequency, low temperature responses of an insulator. [32, 35] The placement of the indices in Eq. (5.1) is not essential for the arguments and calculations in this paper, and the reader can choose to treat α as a Cartesian tensor α_{ij} if desired. [1] As a Cartesian tensor, the traceless part decomposes further into symmetric and antisymmetric parts

$$\tilde{\alpha}_{ij}^S=\frac{1}{2}\left(\tilde{\alpha}_{ij}+\tilde{\alpha}_{ji}\right),\quad \tilde{\alpha}_{ij}^A=\frac{1}{2}\left(\tilde{\alpha}_{ij}-\tilde{\alpha}_{ji}\right)=-\epsilon_{ijk}T_k, \tag{5.3}$$

[1] The index structure can be used as a check, somewhat like dimensional analysis. For example, it is a reminder that $\partial P^x/\partial B^y=\partial M_y/\partial E_x$, which has matching indices, rather than $\partial M_x/\partial E_y$. One only has to remember that $\mathbf{P}$ and $\mathbf{B}$ have upper indices, while $\mathbf{M}$ and $\mathbf{E}$ have lower indices. Vectors with upper indices correspond to directions in space. For example $\mathbf{P}$ has an upper index because it is given by $e\mathbf{r}$, while $\mathbf{E}$ has a lower index, in order for $\mathbf{E}\cdot\mathbf{d}\boldsymbol{\ell}=E_i d\ell^i$ in Faraday's law to be balanced. The index structure is also useful when using "internal coordinates", especially in the case of nonorthorhombic crystals. One writes $\mathbf{P}=\sum_i P^i\mathbf{a}_i$, $\mathbf{B}=\sum_i B^i\mathbf{a}_i$, $\mathbf{M}=\sum_i M_i\mathbf{g}^i/(2\pi)$, and $\mathbf{E}=\sum_i E_i\mathbf{g}^i/(2\pi)$, where $\mathbf{a}_i$ are the primitive lattice vectors and $\mathbf{g}^i$ are the reciprocal lattice vectors, $\mathbf{g}^i\cdot\mathbf{a}_j=2\pi\delta_j^i$. This amounts to setting the primitive vectors to $\hat{\mathbf{x}},\hat{\mathbf{y}},\hat{\mathbf{z}}$.

where $T_i = -\epsilon_{ijk}\tilde{\alpha}_{jk}/2$ is the toroidal response. (Unless otherwise stated, in our work repeated indices are implicitly summed.) The terminology reflects that this part of the orbital magnetoelectric response is related to the "toroidal moment", which is an order parameter that has recently been studied intensively; in a Landau effective free energy, the toroidal moment and the toroidal part of the magnetoelectric response are directly related [19, 6].

The primary goal of this paper is to compute the contribution to α that arises solely from the motion of electrons due to their couplings to the electromagnetic potentials $\rho\phi$ and $-\mathbf{j}\cdot\mathbf{A}$. We call this contribution the orbital magnetoelectric polarizability, or OMP for short. Other effects, such as those mediated by lattice distortions or the Zeeman coupling to the electron's spin, are calculable with known methods [83]. We shall only treat the polarization response to an applied magnetic field here; concurrent work by Malashevich, Souza, Coh, and one of us obtains an equivalent formula by developing methods to compute the orbital magnetization induced by an electrical field [51].

The magnetoelectric tensor's physical consequences arise through the bound current and charge [67, 32], given by $\rho_b = -\mathrm{div}\,\mathbf{P}$ and $\mathbf{J}_b = \partial_t\mathbf{P} + \mathrm{curl}\,\mathbf{M}$. Besides having a ground state value, each moment responds (instantaneously and locally, as appropriate for the low-frequency response of an insulator) to applied electric and magnetic fields, *e.g.*, $P^i = P^i_0 + \chi^{ij}_E E_j + \alpha^i_j B^j$; we will concentrate on the magnetoelectric response. (Unless otherwise stated, in this article repeated indices are implicitly summed.) It is useful to allow α^i_j, a material property, to vary in space and time by allowing the electronic Hamiltonian to vary; this leads to a formula that covers the effects of boundaries and time-dependent shearing of the crystal, for example. Then the relevant terms are

$$\begin{aligned} J^i_b &= (\tilde{\alpha}^l_j\epsilon^{ikj} - \tilde{\alpha}^i_j\epsilon^{jkl})\partial_k E_l + (\partial_t\alpha^i_j)B^j + \epsilon^{ijk}(\partial_j\alpha^l_k)E_l \\ \rho_b &= -\tilde{\alpha}^i_j\partial_i B^j \qquad\qquad\qquad\quad - (\partial_i\alpha^i_j)B^j, \end{aligned} \tag{5.4}$$

We have used two of Maxwell's equations to simplify the first term in each line. The most important point to notice here is that α_θ does not appear except in derivatives, so that any uniform and static contribution to θ has no effect on electrodynamics. Hence in a uniform, static crystal, the components of $\tilde{\alpha}$ can be computed or measured from the current or charge response to spatially varying fields, given by the first term in each line. On the other hand, if we wish similarly to obtain α_θ from charge or current responses to applied fields, we need to consider a crystal that varies either spatially or temporally, so that $\mathbf{E}$ or $\mathbf{B}$ will couple to $\partial_i\alpha_\theta$ or $\partial_t\alpha_\theta$, as in the second terms of Eqs. (5.4). These considerations, which we will elaborate later, motivate our theoretical approach to the OMP in this paper.

We will proceed as follows. In Section II, we present the results of our calculation of the OMP in the independent-electron approximation. This section includes a review of known results, followed by a discussion of the new contributions we compute and when those contributions can be expected to vanish (so that the OMP reduces to the form found in the literature previously). We follow these discussions with a detailed presentation of the calculation in Section III. This calculation involves a novel method for dealing with a uniform magnetic field in a crystal. An alternative derivation is presented in the Appendix.

5.2 General features of the orbital magnetoelectric response

In this section we discuss properties of the OMP and its explicit expression in the independent electron approximation. There is a natural decomposition into two parts, which is, however, not equivalent to the standard symmetry decomposition given in Eq. (5.1) of the Introduction.

The first part is the scalar "Chern-Simons" term α_{CS} obtained by QHZ [67] that contributes only to the trace part α_θ. It is formally similar to the Berry-phase expression for polarization [43] in that it depends only on the wavefunctions, not their energies, which explains the terminology "magneto-electric polarization" introduced by QHZ for α_{CS}. [67] The second part of the response is not simply scalar. It has a different mathematical form that is not built from the Berry connection, looking like a more typical response function in that it involves cross-gap contributions and is not a "moment" determined by the unperturbed wavefunctions. We label this term α_G because of its connection with cross-gap contributions. This term does not seem to have been obtained previously although its physical origin is not complicated.

5.2.1 The OMP expression and the origin of the cross-gap term α_G

We first give the microscopic expression of the new term in the OMP and discuss its interpretation. The later parts of this section explain why the new term vanishes in most of the models that have been introduced in the literature to study the OMP, and discuss to what extent the two terms in the OMP expression are physically separate. The OMP expression that we discuss here will be derived later in Section III as follows: we compute the bulk current in the presence of a small, uniform magnetic field as the crystal Hamiltonian is varied adiabatically. The result is a total time derivative which can be integrated to obtain the magnetically induced bulk polarization.

The most obvious property of the new term α_G in the response is that, unlike the Chern-Simons piece, it has off-diagonal components; for instance, $\partial P^x/\partial B^y \neq 0$. To motivate the expression for α_G intuitively, we note that it is very similar to what one would expect based on simple response theory: An electric dipole moment, $e\mathbf{r}$, is induced when a magnetic field is applied. This field couples linearly to the magnetic dipole moment $(e/4)(\mathbf{r}\times\mathbf{v}-\mathbf{v}\times\mathbf{r})$ (this form takes care of the operator ordering when we go to operators on Bloch states). The expression we actually get for the OMP is expressed in terms of the periodic part of the Bloch wave functions $u_{l\mathbf{k}}$ and the energies $E_{l\mathbf{k}}$ describing the electronic structure of a

crystal:

$$\alpha^i_j = (\alpha_G)^i_j + \alpha_{CS}\delta^i_j \tag{5.5a}$$

$$(\alpha_G)^i_j = \sum_{\substack{n\,\mathrm{occ}\\ m\,\mathrm{unocc}}} \int_{\mathrm{BZ}} \frac{d^3k}{(2\pi)^3} \mathrm{Re}\left\{ \frac{\langle u_{n\mathbf{k}}|e\not{r}^i_{\mathbf{k}}|u_{m\mathbf{k}}\rangle\langle u_{m\mathbf{k}}|e(\mathbf{v}_{\mathbf{k}}\times\not{r}_{\mathbf{k}})_j - e(\not{r}_{\mathbf{k}}\times\mathbf{v}_{\mathbf{k}})_j - 2i\partial H'_{\mathbf{k}}/\partial B^j|u_{n\mathbf{k}}\rangle}{E_{n\mathbf{k}} - E_{m\mathbf{k}}} \right\} \tag{5.5b}$$

$$\alpha_{CS} = -\frac{e^2}{2\hbar}\,\epsilon_{abc} \int_{\mathrm{BZ}} \frac{d^3k}{(2\pi)^3}\,\mathrm{tr}\left[\mathcal{A}^a\partial^b\mathcal{A}^c - \frac{2i}{3}\mathcal{A}^a\mathcal{A}^b\mathcal{A}^c\right]. \tag{5.5c}$$

Here the Berry connection $\mathcal{A}^a_{nn'}(\mathbf{k}) = i\langle u_{n\mathbf{k}}|\partial_{k_a}u_{n'\mathbf{k}}\rangle$ is a matrix on the space of occupied wave functions $u_{n\mathbf{k}}$,[2] and the derivative with an upper index $\partial^a = \partial_{k_a}$ is a k-derivative, as opposed to the spatial derivative ∂_i in Eq. (5.4). The velocity operator is related to the Bloch Hamiltonian, $\hbar v^i(\mathbf{k}) = \partial^i H_{\mathbf{k}}$, while the operator $\not{r}^i_{\mathbf{k}}$ is defined as the derivative $\partial^i\mathcal{P}_{\mathbf{k}}$ of the projection $\mathcal{P}$ onto the occupied bands at $\mathbf{k}$. This operator is closely related to the position operator; its "cross-gap" matrix elements between occupied and unoccupied bands are $\langle u_{m\mathbf{k}}|\not{r}^i_{\mathbf{k}}|u_{n\mathbf{k}}\rangle = \langle u_{n\mathbf{k}}|\not{r}^i_{\mathbf{k}}|u_{m\mathbf{k}}\rangle^* = -i\langle u_{m\mathbf{k}}|r^i|u_{n\mathbf{k}}\rangle$, while its "interior" matrix elements between two occupied bands or two unoccupied bands vanish. Finally, the operator H' is introduced for generality, as discussed in Section 5.3.1; it vanishes for the continuum Schrödinger Hamiltonian and for tight-binding Hamiltonians whose hoppings are all rectilinear, and so will be ignored for most of the analysis that follows. Neglecting this subtlety, the form of α_G is nearly what would be expected for the response in electric dipole moment to a field coupling linearly to the magnetic dipole moment.

The main difference between the explicit form of α_G and the naïve expectation from the dipole moment argument above is that α_G excludes terms of the form $\langle n|\mathbf{r}|m\rangle\langle m|\mathbf{v}|n'\rangle\langle n'|\mathbf{r}|n\rangle$, for example, that include interior matrix elements of $\mathbf{r}$. In some sense, this omission is compensated for by the extra factor of 2 relative to the naïve expectation and by a remainder term, namely, α_{CS}, the Chern-Simons part. The Chern-Simons term α_{CS} alone has appeared previously [67, 20]. The next subsection reviews the properties of α_{CS} and gives a geometrical picture for its discrete ambiguity, which is not present in the α_G term. We then explain how the existence of the previously unreported α_G can be reconciled with previous studies on model Hamiltonians that found only α_{CS}, and then show that the two terms are more intimately related than they first appear.

5.2.2 The Chern-Simons form, axion electrodynamics, and topological insulators

The Chern-Simons response α_{CS} has been discussed at some length in the literature [67, 20]. It does not emerge as clearly as α_G from the intuitive argument above about dipole

[2] In general there will be $\mathbf{k}$ points at which u_n will not be differentiable. Since $\mathcal{A}$ only enters in a matrix trace, however, any orthornormal basis of occupied functions is acceptable, not just the energy basis, and there will be a basis at every point such that $\mathcal{A}$ is well defined. A prescription for a smooth $\mathcal{A}$ can be constructed from maximally localized Wannier functions [52].

moment in a field; rather, in Ref. [20], it was derived by treating the vector potential as a background inhomogeneity and utilizing a general formalism for computing the polarization in such a background [87].

The most important feature of the microscopic expression for the *isotropic* OMP is that it suffers from a discrete ambiguity. The dimensionless parameter θ quantifying the isotropic susceptibility contains the term

$$\theta_{CS} = -\frac{1}{4\pi}\,\epsilon_{abc}\int d^3k\,\mathrm{tr}\left[\mathcal{A}^a\partial^b\mathcal{A}^c - \frac{2i}{3}\mathcal{A}^a\mathcal{A}^b\mathcal{A}^c\right], \tag{5.6}$$

which is only defined up to integer multiples of 2π. This is tied to a "gauge" invariance: ground state properties of a band insulator should only be determined by the ground state density matrix $\rho^g_{\mathbf{k}}$, which is invariant under unitary transformations $U_{nn'}(\mathbf{k})$ that mix the occupied bands. Now, the Berry connection $\mathcal{A}$ is not invariant under such a transformation, but there is no inconsistency because, in the expression for θ_{CS}, all the terms produced by the gauge transformation cancel except for a multiple of 2π. An analogous phenomenon, slightly easier to understand, is found in the case of electric polarization [43]

$$P^i = e\int_{\mathrm{BZ}}\frac{d^3k}{(2\pi)^3}\,\mathcal{A}^i\,, \tag{5.7}$$

which has invariance only up to a discrete "quantum," or ambiguity, which counts the number of times $U(\mathbf{k})$ winds around the Brillouin zone (e.g., if $U_{11} = e^{ik_xa}$ and $U_{ii} = 1$, $i \neq 1$, then P^x changes by one quantum). The Chern-Simons response α_{CS} behaves similarly, although the "winding" that leads to the ambiguity is more complicated (in particular, it is non-Abelian).

These ambiguities can be understood from general arguments, without relying on the explicit formulae. In the case of the polarization, the quantum of uncertainty of P^x, e/S_x, depends on the lattice structure, with S_x the area of a surface unit cell normal to x. The ambiguity results because the bulk polarization does not completely determine the surface charge: isolated surface bands can be filled or emptied, changing the number of surface electrons per cell by an integer. For the magnetoelectric response, the quantum of magnetoelectric polarizability is connected with the fact that θ gives a surface Hall conductance, as can be seen from the term $\mathbf{J}_b = (\boldsymbol{\nabla}\alpha_\theta)\times\mathbf{E}$ in Eq. (5.4). Therefore, the ambiguity in α_θ is just e^2/h, the "quantum of Hall conductance," because it is possible to add a quantum Hall layer to the surface. (This remains a theoretical possibility even if no intrinsic quantum Hall materials have yet been found.)

Now let us show that this ambiguity afflicts only the trace of the susceptibility. This can be seen directly by measuring the bound charge and currents. For example, all the components of $\tilde{\alpha}$ can be deduced from a measurement of ρ_b in the presence of a nonuniform magnetic field [see Eq. (5.4)], but α_θ itself does not determine any bulk properties.

More concretely, one can derive the ambiguities in the magnetoelectric response from the ambiguities in the surface polarization. In a periodic system, which for simplicity we take to have a cubic unit cell, the smallest magnetic field that can be applied without destroying

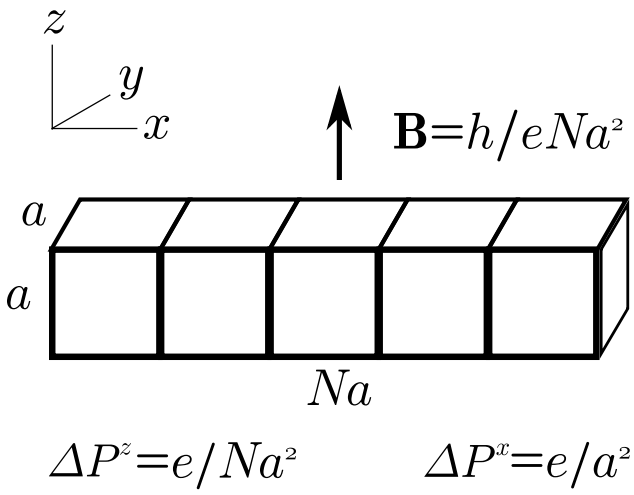

Figure 5.1: A supercell admits a small magnetic flux, and the quantum of polarization transverse to the long direction is correspondingly small, but the quantum for polarization along the long direction is much larger.

the periodicity of the Schrödinger equation corresponds to one flux quantum per unit cell, or $B = h/(eS)$, where S is again a transverse cell area. The ambiguity in the polarization of the system in this magnetic field corresponds to an ambiguity in dP/dB of

$$\frac{\Delta P}{B} = \frac{e/S}{h/(eS)} = \frac{e^2}{h}. \tag{5.8}$$

Hence on purely geometrical grounds there is a natural quantum e^2/h of the *diagonal* magnetoelectric polarizability [20].

In order to see that this uncertainty remains the same when a *small* magnetic field is applied (after all, α is defined as a *linear* response), we will have to construct large supercells in a direction perpendicular to the applied B (Fig. 5.1). While a supercell of N fundamental cells has a less precisely defined polarization (the quantum decreases by a factor N, so the *uncertainty* increases), the minimum field that can be applied also decreases by this factor, so that the uncertainty in the polarizability dP^i/dB^i (no sum) remains constant. On the other hand, if we consider the off-diagonal response, we can consider a supercell with its long axis parallel to the applied B. In this case, the polarization quantum remains constant as the supercell grows large and the minimum applied flux becomes small; the quantum in dP^i/dB^j (for $i \neq j$) then becomes large, which means that the uncertainty vanishes. For this geometry, a small B acts like a continuous parameter, and the change in polarization induced by B can be continuously tracked, even if the absolute polarization remains ambiguous.

Thus, with or without interactions, there is a fundamental difference between the isotropic response and the other components of the response. For the trace-free components, we indeed do not find a quantum of uncertainty in the polarizability formula. In particular, if the toroidal response is defined by $T_i = -\epsilon_{ijk}\tilde{\alpha}_{jk}/2$, then we believe that a "quantum

of toroidal moment" [6] can only exist when there is a spin direction with conserved "up" and "down" densities. (This toroidal moment is typically defined as $\boldsymbol{t} = (1/2)\int \mathbf{r} \times \boldsymbol{\mu} d\mathbf{r}$, with $\boldsymbol{\mu}$ the magnetization density [19], or more generally in terms of a tensor $\mathcal{T}^{ij}$ such that $\partial_i \mathcal{T}^{ij} = -2\mu^j$ [6].) It then reduces to the polarization difference between up and down electrons.

A particular class of materials for which the ambiguity in α_θ is extremely important is the strong topological insulators [22, 56, 72], in which $\theta = \pi$ (Ref. [67]). These are time-reversal (T) symmetric band insulators. At first blush, T invariance should rule out magnetoelectric phenomena at linear order, since $\mathbf{M}$ and $\mathbf{B}$ are T-odd. However, the ambiguity by 2π in θ provides a loophole, since $-\pi$ is equivalent to π. Here we regard the ambiguity/periodicity of θ as a consequence of its microscopic origin (alternately, its coupling to electrons); because θ can be modified by $2\pi n$ by addition of surface integer quantum Hall layers, only θ modulo 2π is a meaningful bulk quantity for systems with charge-e excitations. This is consistent with the gauge-dependence of the integral for α_{CS}. An alternate approach is to derive an ambiguity in θ by assuming that the $U(1)$ fields are derived from a non-Abelian gauge field [82]. The view here that periodicity of θ results from the microscopic coupling to electrons is similar to the conventional understanding of the polarization quantum.

5.2.3 Conditions causing α_G to vanish

It is worthwhile to understand in more detail the conditions under which the response α_G is allowed. It is forbidden in systems with inversion (P) or time-reversal (T) symmetry, which can be seen explicitly from the presence of three k-derivatives acting on gauge-invariant matrices in the formula written in terms of $\mathcal{P}_\mathbf{k}$ and $H_\mathbf{k}$. [3] However, this alone is not sufficient to explain why α_G did not appear in the T-breaking models previously introduced to study the OMP [20, 67, 50]. This is explained by the fact that the interband contribution α_G [Eq. (5.5b)] will also vanish if dispersions satisfy the following "degeneracy" and "reflection" conditions:

- At a given $\mathbf{k}$, all the occupied valence bands have the same energy $E^{\mathrm{v}}_\mathbf{k}$.
- Similarly, all the unoccupied conduction bands have the same energy $E^{\mathrm{c}}_\mathbf{k}$.
- $E^{\mathrm{v}}_\mathbf{k} + E^{\mathrm{c}}_\mathbf{k}$ is independent of $\mathbf{k}$ (and can be taken to be zero).

[3] This argument does not quite hold for the Chern-Simons piece since the gauge chosen for the Berry connection may not share the symmetry of the system, accounting for the nontrivial value in a topological insulator.

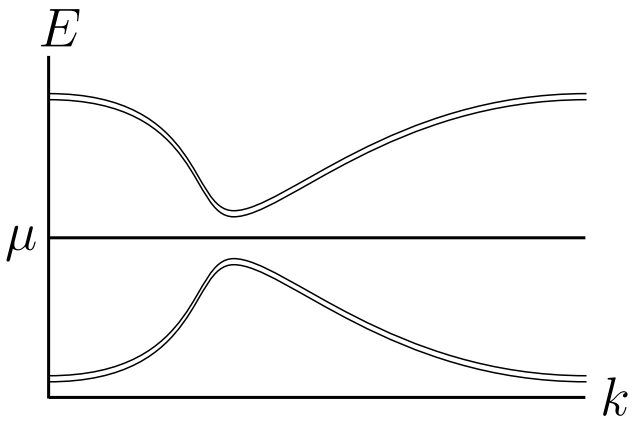

Figure 5.2: Schematic band structure that leads to vanishing α_G. The bands below the chemical potential are degenerate with energy $E^{\rm v}_{\mathbf{k}}$, while the bands above the chemical potential have energy $E^{\rm c}_{\mathbf{k}} = \text{const.} - E^{\rm v}_{\mathbf{k}}$.

This can be seen immediately in an expanded form of the integrand of α_G, [see Eqs. (B.11c) and (B.11d)]

$$-\sum_{\substack{n,n'\,\text{occ}\\ m\,\text{unocc}}}(E_n-E_{n'})\frac{\langle\partial^b n|n'\rangle\langle\partial^a n'|m\rangle\langle m|\partial^i n\rangle}{E_n-E_m}+\sum_{\substack{n\,\text{occ}\\ m,m'\,\text{unocc}}}(E_m-E_{m'})\frac{\langle\partial^b n|m'\rangle\langle m'|\partial^a m\rangle\langle m|\partial^i n\rangle}{E_n-E_m}$$
$$-\sum_{\substack{n\,\text{occ}\\ m\,\text{unocc}}}\partial^b(E_n+E_m)\frac{\langle\partial^a n|m\rangle\langle m|\partial^i n\rangle}{E_n-E_m}, \quad (5.9)$$

where $|n\rangle = |u_{n\mathbf{k}}\rangle$ and $E_n = E_{n\mathbf{k}}$, etc. Such a structure is automatic when only two orbitals (with both spin states) are taken into account and the system has particle-hole and PT symmetries. PT symmetry guarantees that the bands remain spin-degenerate even if spin is not a good quantum number. To see this, recall that T acts on wave functions as $i\sigma^y K$ and maps $\mathbf{k} \to -\mathbf{k}$. Here, K is complex conjugation and σ^y takes the form of the usual Pauli matrix in the z basis of spin. Then P maps $\mathbf{k} \to -\mathbf{k}$ again, so that PT effectively acts as "T at each $\mathbf{k}$" [3] . Then particle-hole symmetry implies that the dispersion is reflection-symmetric, $E^{\rm v}_{\mathbf{k}} = -E^{\rm c}_{\mathbf{k}}$.

Most model Hamiltonians discussed in the literature that access the topological insulator phase [22, 67, 20, 36, 50], as well as the Dirac Hamiltonian (in the context of which the axion electrodynamics was first discussed [82]), can be defined in terms of a Clifford algebra,[4] and this ensures that the dispersions are degenerate and reflection symmetric. The only exception of which we are aware is the model of Guo and Franz on the pyrochlore lattice,

[4]Concretely, the generators of a Clifford algebra are a set of N_c matrices Γ_a that satisfy the relation $\{\Gamma_a, \Gamma_b\} = 2\delta_{ab}$. Then the Hamiltonians cited take the form $H(\mathbf{k}) = \sum_{a=1}^{N_c} \epsilon_a(\mathbf{k})\Gamma_a$. This automatically satisfies the degeneracy and dispersion-reflection properties, since $H(\mathbf{k})$'s eigenvalues are $\pm\sqrt{\sum_a \epsilon_a(\mathbf{k})^2}$.

which has four orbitals per unit cell [28]. The topological insulator phase itself will not have a contribution from α_G, since it is T-invariant, and so the Guo and Franz model will not show such a response; however, the addition of any T-breaking perturbation to their model should produce off-diagonal magnetoelectric responses.

Finally, there is a simple mathematical condition that will cause α_G to vanish. Namely, α_G decreases as the gap becomes large without changing the wave functions, and in the limit of infinite bulk gap the only magnetoelectric response comes from the Chern-Simons part, which is not sensitive to the energies and depends only on the electron wave functions.

5.2.4 Is the Chern-Simons contribution physically distinct?

Apart from the ambiguity in α_{CS} that is not present in α_G, there seems to be no real physical distinction between the two terms of the linear magnetoelectric response. We discuss two aspects that relate to this observation below.

Localized vs. itinerant contributions

The ambiguity in θ_{CS} can be interpreted as a manifestation of the fact that bulk quantities cannot determine the surface quantum Hall conductance, since a two-dimensional quantum Hall layer could appear on a surface independent of bulk properties. This suggests, perhaps, that the Chern-Simons term appears only in bulk systems with extended wave functions, and is a consequence of the itinerant electrons, while α_G is a localized molecule-like contribution. However, this turns out not to be the case.

Consider a periodic array of isolated molecules, which is an extreme limit of the class of crystalline insulators. Such a system has flat bands, with energies equal to the energies of the molecular states, since the electrons cannot propagate. It is certainly possible to construct a molecular system where all the unoccupied states have one energy and all the occupied states have another, by tuning the potentials. In this case α_G will vanish. However, such a molecule can still display a magnetoelectric response; it will therefore have to be given by α_{CS} (and so restricted to diagonal responses). For example, consider the "molecule" of Fig. (5.3) with the shape of a regular tetrahedron. If the two low-energy levels are occupied, the magnetoelectric response is

$$\frac{\partial P^i}{\partial B^j} = \pm\delta^i_j \frac{1}{\sqrt{6}} \frac{e^2}{\hbar},$$

where P^i here is the electric dipole moment divided by the volume of the tetrahedron; the sign of the polarizability reverses when the complex phases are reversed. This shows that the Chern-Simons term does not require delocalized orbitals.

Additivity

Another argument against distinguishing between the Chern-Simons part and the rest of the susceptibility is based on band additivity. When interactions are not taken into account, each occupied band can be regarded as an independent physical system (at least if there are no band crossings). Applying a magnetic field causes each band n to become polarized by a certain amount $\mathbf{P}_n$, and so the net polarization should be $\mathbf{P} = \sum_n \mathbf{P}_n$. The Pauli exclusion

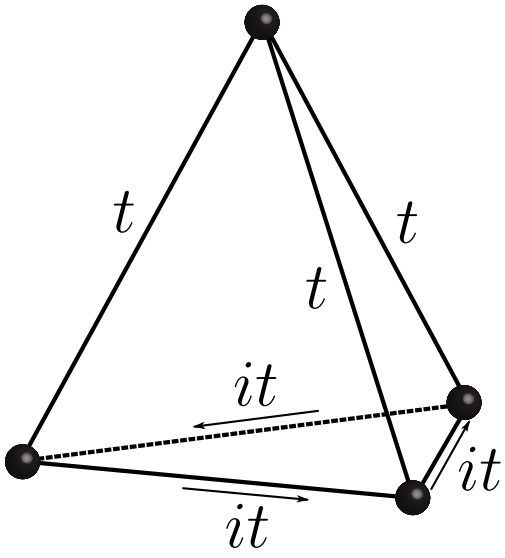

Figure 5.3: A tetrahedral tight-binding molecule for spinless electrons, with one orbital per site and complex hoppings. The hopping integrals are all equal, except that those around one face have a phase of i relative to the other three. There are then two pairs of degenerate levels.

principle does not lead to any "interactions" between pairs of bands, because the polarization (like any single-body operator) can be written as the sum of the mean polarization in each of the orthonormal occupied states.

Now the Chern-Simons form does not look particularly additive in this sense, and is not by itself. Because it is the trace of a matrix product in the occupied subspace, it necessarily involves matrix elements between different occupied states, while an additive formula would not. Nevertheless, α_G and α_{CS} are together additive, as can be seen most simply in Eq. (B.9), where the two terms combine into a single sum over occupied bands. In terms of α_G and α_{CS} separately, one finds that when the values of α_G, assuming just band 1 or 2 is occupied, are added together, some terms occur that are not present in the expression for $\alpha_G(1+2)$ (where both bands are occupied), and vice-versa. Using Eqs. (B.11c) and (B.11d) we see, in fact, that α_G is a sum of contributions which depend on three bands, as $\alpha_G = \sum_{n,m,m'} C(n;m,m') + \sum_{n,n',m} D(n,n';m)$. Terms such as $C(1;2,m')$ are not present in the expression for $\alpha_G(1+2)$. (Likewise $D(1,2;m)$ appears in $\alpha_G(1+2)$ but not in $\alpha_G(1)$ and $\alpha_G(2)$.) Adding up the discrepancies, one finds that the energy-denominators all cancel, and the non-diagonal terms from the Chern-Simons form appear!

Seemingly paradoxical is the fact that for band structures satisfying the degeneracy and reflection conditions of the last subsection, the magnetoelectric susceptibility is given by the Chern-Simons term alone, which does not seem to be additive. However, the additivity property applies only to bands that do not cross. It does not make any sense to ask whether the susceptibility is the sum over the susceptibilities for the systems in which just one of the

degenerate bands is occupied, since those systems are not gapped.

5.3 The OMP as the current in response to a chemical change

Now we will tackle the problem of deriving the formula for the OMP α discussed in the last section. There are two impediments we need to overcome, a physical one, and a more technical one (which we will overcome starting from an insight of Levinson) [49].

In order to determine α, we would like to carry out a thought experiment in which a crystal is exposed to appropriate electromagnetic fields. For specificity, we will apply a uniform magnetic field. To make the calculation of the response clean, we wish to deal with an infinite crystal. Then the polarization does not simply reduce to the first moment of the charge density [43], so we will instead have to calculate the current or charge distribution induced by the fields, and then use Eq. (5.4) to deduce α. If both the crystal and the electromagnetic fields are independent of space and time, there is no macroscopic charge or current density. We will assume *spatial* uniformity, so that there are two choices for how to proceed. Either the magnetic field can be varied in time or the crystal parameters, and thus α, can be varied. In either case, we measure the current that flows through the bulk and try to determine α. As ever, the diagonal response α_θ is the most difficult to capture: while either time-dependent experiment can be used to determine $\tilde{\alpha}$, only the latter approach sheds light on the value of α_θ.

To see why α_θ can be determined only in this way (given that we want to work with a spatially homogeneous geometry), let us discuss how currents flow through the crystal. The necessity of varying the crystal in time can be deduced from Maxwell's equations (see below) but we will give a more intuitive discussion here. Suppose that $\tilde{\alpha} = 0$. Then in an applied magnetic field there is a polarization $\mathbf{P} = \alpha_\theta \mathbf{B}$; thus the crystal gets charged at the surface. As the magnetic field is turned on, this surface charge has to build up (charge density $\hat{\mathbf{n}} \cdot \mathbf{P}$). This occurs entirely due to flows of charge *along the surface*. Suppose, for example, that the sample is a cylinder (radius R) with the magnetic field along its z axis, as illustrated in Fig. 5.4(a). Then an electric field $\mathbf{E}_{\text{ind}} = -\dot{B}R\hat{\boldsymbol{\phi}}/2$ is induced at the surface according to Faraday's law. Besides being the magnetoelectric response, θ also represents the Hall coefficient for surface currents. Therefore, a current of $J_s = \alpha_\theta \dot{B} R/2$ flows to the top of the cylinder, adding up to a surface charge of $2\pi R \int J_s(t) dt = \alpha_\theta B_f \pi R^2$ and producing the entire polarization $\alpha_\theta B_f$. No current flows through the bulk! In fact, the Hall conductance on the circular face produces a radial current as well, so that the charge distributes over the surface rather than just accumulating in a ring. Note that the surface current grows with the radius of the cylinder. This sounds like a nonlocal response, but it can be understood as follows: the electric field is determined by the non-local Faraday law, but the *crystal's* response to the electric field (namely, the surface current) is local.

The current distribution can be understood directly from Maxwell's equations: there are two contributions to the bulk current, $\partial_t \mathbf{P}$ and curl $\mathbf{M}$. The polarization is $\alpha_\theta \mathbf{B}$ while

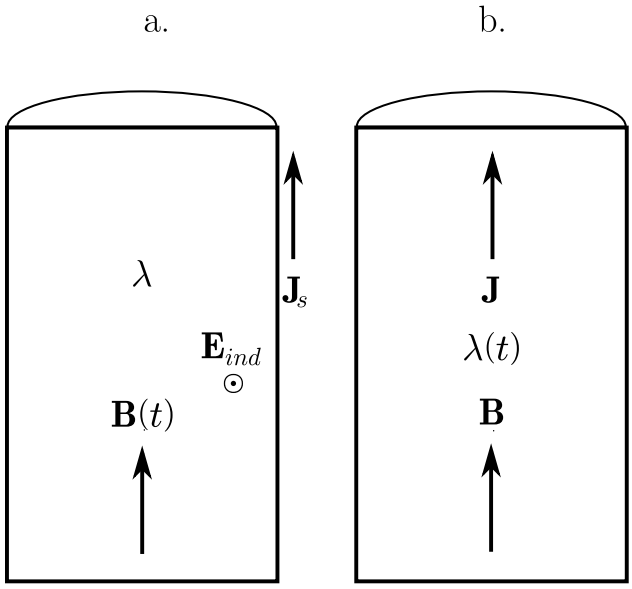

Figure 5.4: As outlined in the text, (a) turning on a magnetic field produces a macroscopic polarization through the flow of surface currents, while (b) varying the crystal Hamiltonian in the presence of a fixed magnetic field produces a polarization through the flow of current through the bulk.

the magnetization is indirectly produced by the induced electric field, $\alpha_\theta \mathbf{E}_{\rm ind}$. The two contributions thus cancel by Faraday's law in the bulk: $\mathbf{J}_b^{\rm bulk} = \alpha_\theta(\partial_t \mathbf{B} + \text{curl } \mathbf{E}) = 0$. There *is* a surface current because α_θ is discontinuous there.

On the other hand, if θ changes in time, while the magnetic field is time-independent (as in Fig. 5.4b), the polarization at the ends of the cylinder builds up entirely by means of flows of charge *through the bulk*. Surface flows cannot be large enough to explain the net polarization in this situation. Since there is no induced electric field, the surface current is just proportional to the lateral surface area and is negligible compared to the bulk current. Therefore the bulk current is equal to $(\partial_t \alpha_\theta)B$ and can be integrated to give $\alpha_\theta B$.

For the other component of the OMP, $\tilde{\alpha}$, either thought experiment can be used. The simplest approach, however, is still the crystal-variation method, since the surface currents are negligible in that case, [5] and in any case this method allows us to find all the components of α simultaneously.

Difficulties with the operator $\mathbf{r}$ *and uniform magnetic fields*

There are two technical difficulties in the theory. First, the operator $\mathbf{r}$ has unbounded matrix elements and thus the matrix elements of the magnetic dipole moment $(e/4)(\mathbf{v} \times \mathbf{r} - \mathbf{r} \times \mathbf{v})$ are not well-defined. This rules out the straightforward use of perturbation theory to calculate the electric dipole moment of an infinite crystal in a uniform magnetic field.

[5] The surface currents and bulk currents both give a definite contribution to the polarization when the magnetic field is varied.

Second, if we consider a crystal in a uniform magnetic field, Bloch's theorem does not hold. Although the magnetic field is uniform, the vector potential that appears in the Hamiltonian depends on $\mathbf{r}$.

We avoid the problems of $\mathbf{r}$ as follows. The key idea is to work with the ground state density matrix, rather than wave functions. The individual eigenstates change drastically when a magnetic field, no matter how small, is applied (consider the difference between a plane wave and a localized Landau level). However, the *density matrix* of an insulator summed over the occupied bands only changes by a small amount when $\mathbf{B}$ is applied; over short distances the magnetic field cannot have a strong effect (even in the example of Landau levels), and the density matrix has only short-range correlations because it describes an insulating state. More technically, we show (subsection 5.3.1) that the broken translation invariance of any single-body operator $\mathcal{O}$ (such as the density matrix) can be dealt with by factoring out an Aharonov-Bohm-like phase from its matrix $\mathcal{O}_{\mathbf{r}_1\mathbf{r}_2}$. This solves the problem of the nonuniform gauge field and leads to expressions that depend only on differences between $\mathbf{r}$'s. In addition, since the exponentially decaying ground-state density matrix appears multiplying every expression, the factors of $\mathbf{r}_1 - \mathbf{r}_2$ are suppressed.

The calculation then proceeds as follows. First, using the symmetries of the electron Hamiltonian in a uniform magnetic field, we find how the density matrix changes in a weak magnetic field. Next we compute the current response to an adiabatic variation of the crystal Hamiltonian. Finally, we show that this current can be expressed as a total time derivative, and therefore can be integrated to give the polarization; at linear order in $\mathbf{B}$ we can read off the coefficients, the magnetoelectric tensor α.

5.3.1 Single-body operators for a uniform magnetic field

Recall the form of the Schrödinger Hamiltonian for a single electron in a crystal and under the influence of a magnetic field,

$$H_S(\mathbf{p}, \mathbf{r}) = \frac{1}{2m} \left[\mathbf{p} - e\mathbf{A}(\mathbf{r})\right]^2 + V(\mathbf{r}), \tag{5.10}$$

where $V(\mathbf{r} + \mathbf{R}) = V(\mathbf{r})$ for lattice vectors $\mathbf{R}$. The necessity of using the vector potential $\mathbf{A}$ seems at first to spoil the lattice translation symmetry one would expect in a uniform magnetic field. However, as noted by Brown [10] and Zak [93], a more subtle form of translation symmetry remains. In particular, choosing the gauge

$$\mathbf{A} = \frac{1}{2}\mathbf{B} \times \mathbf{r}, \tag{5.11}$$

the Hamiltonian has "magnetic translation symmetry":

$$H_S(\mathbf{p}, \mathbf{r} + \mathbf{R}) = e^{ie\mathbf{B}\cdot(\mathbf{R}\times\mathbf{r})/2\hbar} H_S(\mathbf{p}, \mathbf{r}) e^{-ie\mathbf{B}\cdot(\mathbf{R}\times\mathbf{r})/2\hbar}. \tag{5.12}$$

This condition defines magnetic translation symmetry for general single-body operators. Any operator $\mathcal{O}$ possessing this symmetry can be written in the position basis as

$$\mathcal{O}_{\mathbf{r}_1\mathbf{r}_2} = \bar{\mathcal{O}}_{\mathbf{r}_1\mathbf{r}_2} e^{-ie\mathbf{B}\cdot(\mathbf{r}_1\times\mathbf{r}_2)/2\hbar}, \tag{5.13a}$$

where $\bar{\mathcal{O}}$ has lattice translation invariance,

$$\bar{\mathcal{O}}_{\mathbf{r}_1+\mathbf{R},\mathbf{r}_2+\mathbf{R}} = \bar{\mathcal{O}}_{\mathbf{r}_1\mathbf{r}_2}. \tag{5.13b}$$

Note that the phase is just $(ie/\hbar)\int d\boldsymbol{\ell}\cdot\mathbf{A}$ calculated along the straight line from $\mathbf{r}_2$ to $\mathbf{r}_1$, which agrees with the intuition that comes from writing the second-quantized form of the operator,

$$\mathcal{O} = \int d^3r_1 d^3r_2\, \bar{\mathcal{O}}_{\mathbf{r}_1\mathbf{r}_2} c^\dagger_{\mathbf{r}_1} e^{-ie\mathbf{B}\cdot(\mathbf{r}_1\times\mathbf{r}_2)/2\hbar} c_{\mathbf{r}_2}. \tag{5.14}$$

This argument shows how to couple general Hamiltonians to uniform fields:

$$H = \exp[(ie/\hbar)\int_{\mathbf{r}_2}^{\mathbf{r}_1} d\boldsymbol{\ell}\cdot\mathbf{A}][H_{0\mathbf{r}_1\mathbf{r}_2} + H'_{\mathbf{r}_1\mathbf{r}_2}(B)].$$

The vector potential appears explicitly only in $\mathbf{A}$, while $H'(B)$ gives the rest of the dependence on the magnetic field. The Schrödinger Hamiltonian (5.10) is obtained if we take

$$\bar{H}_{0,\mathbf{r}_1\mathbf{r}_2} = \left[-\frac{\hbar^2}{2m}\nabla^2_{\mathbf{r}_2} + V(\mathbf{r}_2)\right]\delta^{(3)}(\mathbf{r}_2-\mathbf{r}_1). \tag{5.15}$$

and set $H'=0$. Our results also apply to tight-binding models. We introduce H' to capture the possibility that in a tight-binding model the hoppings will not be rectilinear, and hence that the phases in Eq. (5.14) do not capture the full field dependence of the Hamiltonian.

5.3.2 The ground state density operator

We find it convenient to work with the one-body density matrix ρ^g, or equivalently the projector onto the occupied states, whenever possible, because it is a basis-independent object. Also, in an insulator, $\rho^g_{\mathbf{r}_1\mathbf{r}_2}$ is exponentially suppressed in the distance $|\mathbf{r}_2-\mathbf{r}_1|$, which tempers the divergences that arise from the unboundedness of $\mathbf{r}$ [45]. In any case, if the ground state is translationally symmetric, the structure described above will apply to ρ_g and we can be sure that *the density matrix has translational symmetry apart from a phase*:

$$\rho^g_{\mathbf{r}_1\mathbf{r}_2} = \bar{\rho}^g_{\mathbf{r}_1\mathbf{r}_2} e^{-ie\mathbf{B}\cdot(\mathbf{r}_1\times\mathbf{r}_2)/2\hbar}, \tag{5.16}$$

where $\bar{\rho}^g$ possesses the translation symmetry of the crystal lattice and hence should connect smoothly to the field-free density matrix. Hence we will write

$$\bar{\rho}^g = \rho_0 + \rho', \tag{5.17}$$

where ρ_0 is the density operator of the crystal in the absence of the

Density matrix perturbation theory: Now we have to calculate ρ', using a kind of perturbation theory that focuses on density matrices rather than wave functions, since the wavefunctions suffer from the problems discussed above. This perturbation theory starts from

two characteristic properties of the density matrix: it commutes with H, and for fermions at zero temperature it is a projection operator. The latter means that all states are either occupied or unoccupied, so the eigenvalues of the density operator are 0 and 1, which is formalized as

$$\rho^g \rho^g = \rho^g \tag{5.18}$$

(idempotency) [17]. Expressed in the position basis,

$$\begin{aligned} \rho^g_{\mathbf{r}_1\mathbf{r}_3} &= \int d\mathbf{r}_2 \rho^g_{\mathbf{r}_1\mathbf{r}_2}\rho^g_{\mathbf{r}_2\mathbf{r}_3} \\ \bar{\rho}^g_{\mathbf{r}_1\mathbf{r}_3} &= \int d\mathbf{r}_2 \bar{\rho}^g_{\mathbf{r}_1\mathbf{r}_2}\bar{\rho}^g_{\mathbf{r}_2\mathbf{r}_3} e^{-(ie/2\hbar)\mathbf{B}\cdot(\mathbf{r}_1\times\mathbf{r}_2+\mathbf{r}_2\times\mathbf{r}_3+\mathbf{r}_3\times\mathbf{r}_1)}. \end{aligned} \tag{5.19}$$

The exponent is just $-i\phi_{123}/\phi_0$, proportional to the magnetic flux through triangle 123, and the exponential can be expanded for small B. At first order this gives

$$\rho'_{\mathbf{r}_1\mathbf{r}_3} = \int d\mathbf{r}_2 [\rho'_{\mathbf{r}_1\mathbf{r}_2}\rho_{0\mathbf{r}_2\mathbf{r}_3} + \rho_{0\mathbf{r}_1\mathbf{r}_2}\rho'_{\mathbf{r}_2\mathbf{r}_3} - \rho_{0\mathbf{r}_1\mathbf{r}_2}\rho_{0\mathbf{r}_2\mathbf{r}_3}\,(i\phi_{123}/\phi_0)]. \tag{5.20}$$

The problem of the unbounded $\mathbf{r}$'s *is resolved* in this equation because the area A_{123} of the triangle is finite and independent of the origin, and also suppressed by the factor of ρ.

Calculation of ρ': In the last term of Eq. (5.20), we can rewrite $2A_{123} = \mathbf{r}_1 \times \mathbf{r}_2 + \mathbf{r}_2 \times \mathbf{r}_3 + \mathbf{r}_3 \times \mathbf{r}_1 = (\mathbf{r}_2 - \mathbf{r}_1) \times (\mathbf{r}_3 - \mathbf{r}_2)$ and then use $(\mathbf{r}_2 - \mathbf{r}_1)\rho_{0\mathbf{r}_1\mathbf{r}_2} = [\rho_0, \mathbf{r}]_{\mathbf{r}_1\mathbf{r}_2}$, etc., to obtain

$$(1-\rho_0)\rho'(1-\rho_0) - \rho_0\rho'\rho_0 = -i\frac{e}{2\hbar}\mathbf{B}\cdot([\rho_0,\mathbf{r}]\times[\rho_0,\mathbf{r}]). \tag{5.21}$$

If we define

$$\bar{H} = H_0 + H', \tag{5.22}$$

then analogous manipulations (including $(\mathbf{r}_1 - \mathbf{r}_2)H_{\mathbf{r}_1\mathbf{r}_2} = i\hbar\mathbf{v}_{\mathbf{r}_1\mathbf{r}_2}$) on the equation $[H, \rho^g] = 0$ give

$$[\rho', H_0] = \frac{e}{2}\mathbf{B}\cdot([\rho_0,\mathbf{r}]\times\mathbf{v} - \mathbf{v}\times[\rho_0,\mathbf{r}]) - [\rho_0, H']. \tag{5.23}$$

Eqs. (5.21) and (5.23) have an intuitive meaning. The former equation determines the "interior" matrix elements of ρ', those between two occupied or two unoccupied states of the zero-field Hamiltonian. A perturbation with the full crystal symmetry does not change the interior matrix elements of the density matrix because of the exclusion principle [53]. In our case, however, multiplying ρ_0 by the phase $e^{(ie/2\hbar)\mathbf{B}\cdot\mathbf{r}_1\times\mathbf{r}_2}$ gives a density matrix with the correct magnetic translation symmetry, but also changes the momentum of the states and so results in a small probability for states to be douby occupied. Therefore ρ' must correct for this "violation of the exclusion principle." On the other hand Eq. (5.23) determines the "cross-gap" matrix elements of ρ' (those between unoccupied and occupied states). These matrix elements capture the expected "transitions across the gap" induced by the field. The rest of this section is devoted to calculating all these matrix elements. The results are given in Eqs. (5.24) and (5.28); the derivations could be skipped on a first reading.

Calculation of ρ'. Precisely speaking, Eq. (5.21) gives the matrix elements of $\bar{\rho}^g$ between pairs of occupied (n and n') or unoccupied (m and m') states:

$$\begin{aligned}\langle\psi_{n\mathbf{k}}|\bar{\rho}^g|\psi_{n'\mathbf{k}}\rangle &= \delta_{nn'} - \frac{e}{4\hbar}B^j\epsilon_{jab}\mathcal{F}^{ab}_{nn'}(\mathbf{k}) \\ \langle\psi_{m\mathbf{k}}|(1-\bar{\rho}^g)|\psi_{m'\mathbf{k}}\rangle &= \delta_{mm'} - \frac{e}{4\hbar}B^j\epsilon_{jab}\check{\mathcal{F}}^{ab}_{mm'}(\mathbf{k}),\end{aligned} \tag{5.24}$$

where $\mathcal{F}$ is the non-Abelian Berry curvature associated with the occupied bands,

$$\begin{aligned}\mathcal{F}^{ab}_{nn'} &= i\langle u_{n\mathbf{k}}|\partial^a\mathcal{P}_{\mathbf{k}}\partial^b\mathcal{P}_{\mathbf{k}} - \partial^b\mathcal{P}_{\mathbf{k}}\partial^a\mathcal{P}_{\mathbf{k}}|u_{n'\mathbf{k}}\rangle \\ &= \partial^a\mathcal{A}^b_{nn'} - \partial^b\mathcal{A}^a_{nn'} - i[\mathcal{A}^a,\mathcal{A}^b]_{nn'},\end{aligned} \tag{5.25}$$

and $\check{\mathcal{F}}$ is the corresponding quantity for the unoccupied bands. To derive these relations, we use

$$\rho^g = \int_{\mathrm{BZ}}\frac{d^3k}{(2\pi)^3}e^{i\mathbf{k}\cdot\mathbf{r}}\mathcal{P}_{\mathbf{k}}e^{-i\mathbf{k}\cdot\mathbf{r}} \tag{5.26}$$

where $\mathcal{P} = \sum_{n\,\mathrm{occ}}|u_{n\mathbf{k}}\rangle\langle u_{n\mathbf{k}}|$ is the projector onto filled bands at $\mathbf{k}$. This gives

$$\begin{aligned}i[\rho^g,\mathbf{r}] &= \int_{\mathrm{BZ}}\frac{d^3k}{(2\pi)^3}e^{i\mathbf{k}\cdot\mathbf{r}}(\boldsymbol{\nabla}_{\mathbf{k}}\mathcal{P}_{\mathbf{k}})e^{-i\mathbf{k}\cdot\mathbf{r}} \\ &= \int_{\mathrm{BZ}}\frac{d^3k}{(2\pi)^3}e^{i\mathbf{k}\cdot\mathbf{r}}\ \not{r}_{\mathbf{k}}e^{-i\mathbf{k}\cdot\mathbf{r}}\end{aligned} \tag{5.27}$$

after discarding a total derivative. The notation $\not{r} = \boldsymbol{\nabla}_{\mathbf{k}}\mathcal{P}$ was introduced in Eq. (5.5).

By contrast, Eq. (5.23) describes to what extent $\bar{\rho}^g$ fails to commute with H_0, the crystal Hamiltonian, and gives the matrix elements of ρ' between occupied and unoccupied states. In this sense it is analogous to the more usual results for density-matrix perturbation theory [53]. In the basis of unperturbed energy eigenstates,

$$\langle\psi_{n\mathbf{k}}|\rho'|\psi_{m\mathbf{k}}\rangle = i\frac{e}{2\hbar}B^j\epsilon_{jab}\frac{\langle u_{n\mathbf{k}}|\{\partial^a\mathcal{P}_{\mathbf{k}},\partial^b H_{\mathbf{k}}\}|u_{m\mathbf{k}}\rangle}{E_{n\mathbf{k}}-E_{m\mathbf{k}}} + \frac{\langle u_{n\mathbf{k}}|H'_{\mathbf{k}}|u_{m\mathbf{k}}\rangle}{E_{n\mathbf{k}}-E_{m\mathbf{k}}}. \tag{5.28}$$

Recall that $\hbar v^b = \partial^b H_{\mathbf{k}}$ and that H' is introduced only to capture unusual situations such as tight-binding models with non-straight hoppings, and vanishes for the continuum problem. Eqs. (5.24) and (5.28) are the key technical results of this formalism, good to linear order in the magnetic field.

5.3.3 Adiabatic current

Now we need to calculate the current as the Hamiltonian is changed slowly as a function of time, as in the ordinary theory of polarization [70, 43]. We have to be careful, however, since the current vanishes in the zero-order adiabatic ground state described by density matrix $\rho^g(t)$. It is necessary to go to first order in adiabatic perturbation theory, which takes

account of the fact that the true dynamical density matrix $\rho(t)$ has an extra contribution proportional to $dH/dt = \dot{H}$. However, once the current has been expressed in terms of $\dot{\rho}$, which is proportional to $\dot{H}$, the distinction is no longer important and the adiabatic approximation can be made.

Preparing for the adiabatic approximation: We can write the current as

$$\mathbf{J}(t) = \frac{e}{\Omega}\mathrm{Tr}\rho(t)\mathbf{v} = \frac{e}{\Omega}\frac{i}{\hbar}\mathrm{Tr}\rho[H,\mathbf{r}] \tag{5.29}$$

where Ω is the crystal volume. Here H is the full Hamiltonian including the magnetic field. By unitarity of time evolution, it remains a projector if the initial state describes filled bands only. The operator $\mathbf{r}$ appears here, but in a commutator. Since $\langle\mathbf{r}_1|[\mathcal{O},\mathbf{r}]|\mathbf{r}_2\rangle = (\mathbf{r}_2 - \mathbf{r}_1)\mathcal{O}_{\mathbf{r}_1\mathbf{r}_2}$, such expressions do not suffer from the difficulties of an "unprotected" $\mathbf{r}$, namely its unboundedness. We can only use cyclicity of the trace to the extent that this property can be preserved. In particular, the expression $\mathrm{Tr}\,\mathbf{r}[\rho,H]$, which seems formally equivalent to Eq. (5.29), poses problems, but

$$\mathbf{J}(t) = \frac{e}{\Omega}\frac{i}{\hbar}\mathrm{Tr}[\rho,[\rho,\mathbf{r}]][\rho,H] \tag{5.30}$$

does not. This expression can be derived from Eq. (5.29) using again the idempotency of ρ ($\rho\rho = \rho$). Using the equation of motion for the density matrix,

$$i\hbar\dot{\rho}(t) = [H(t),\rho(t)], \tag{5.31}$$

and making the approximation $\rho \approx \rho^g$ on the right-hand side at this stage, the current becomes

$$\mathbf{J} = \frac{e}{\Omega}\int d\mathbf{r}_1 d\mathbf{r}_2 d\mathbf{r}_3(\mathbf{r}_1 - 2\mathbf{r}_2 + \mathbf{r}_3)\rho^g_{\mathbf{r}_1\mathbf{r}_2}\rho^g_{\mathbf{r}_2\mathbf{r}_3}\dot{\rho}^g_{\mathbf{r}_3\mathbf{r}_1}. \tag{5.32}$$

Magnetic field dependence of the current: The considerations given in the last subsection make the integrand

$$\rho^g_{\mathbf{r}_1\mathbf{r}_2}\rho^g_{\mathbf{r}_2\mathbf{r}_3}\dot{\rho}^g_{\mathbf{r}_3\mathbf{r}_1} = \bar{\rho}^g_{\mathbf{r}_1\mathbf{r}_2}\bar{\rho}^g_{\mathbf{r}_2\mathbf{r}_3}\dot{\bar{\rho}}^g_{\mathbf{r}_3\mathbf{r}_1}e^{-i\phi_{123}/\phi_0}, \tag{5.33}$$

where, again, $\phi_{123} = \mathbf{B}\cdot(\mathbf{r}_1\times\mathbf{r}_2 + \mathbf{r}_2\times\mathbf{r}_3 + \mathbf{r}_3\times\mathbf{r}_1)/2$ is the magnetic flux through the triangle with vertices $\mathbf{r}_1\mathbf{r}_2\mathbf{r}_3$ and does not suffer from the pathologies of $\mathbf{r}$ itself, which allows us to expand $e^{-i\phi_{123}/\phi_0} = 1 - i\phi_{123}/\phi_0$ to lowest order in B (recall again that the matrix elements of ρ are exponentially suppressed with distances).

Recalling our division $\bar{\rho}^g = \rho_0 + \rho'$ where ρ' is of first order in the magnetic field, Eq. (5.32) becomes

$$\mathbf{J} = \frac{e}{\Omega}\int d\mathbf{r}_1 d\mathbf{r}_2 d\mathbf{r}_3(\mathbf{r}_1 - 2\mathbf{r}_2 + \mathbf{r}_3)\Big[\rho_{0\mathbf{r}_1\mathbf{r}_2}\rho_{0\mathbf{r}_2\mathbf{r}_3}\dot{\rho}'_{\mathbf{r}_3\mathbf{r}_1} + \rho'_{\mathbf{r}_1\mathbf{r}_2}\rho_{0\mathbf{r}_2\mathbf{r}_3}\dot{\rho}_{0\mathbf{r}_3\mathbf{r}_1} + \rho_{0\mathbf{r}_1\mathbf{r}_2}\rho'_{\mathbf{r}_2\mathbf{r}_3}\dot{\rho}_{0\mathbf{r}_3\mathbf{r}_1} - i\frac{\phi_{123}}{\phi_0}\rho_{0\mathbf{r}_1\mathbf{r}_2}\rho_{0\mathbf{r}_2\mathbf{r}_3}\dot{\rho}_{0\mathbf{r}_3\mathbf{r}_1}\Big] \tag{5.34}$$

at first order. The rest of the calculation involves substituting the expressions for the magnetic-field dependence of ρ^g obtained earlier, and integrating the result to obtain α. The energy-dependent part of α, namely α_G, will come from the mixing of the occupied and unoccupied bands, Eq. (5.28). The Chern-Simons term will come from the "exclusion-principle–correcting" terms, Eq. (5.24), as well as the ϕ_{123} term in the previous equation.

Integrating the results: The four terms in the current can be collected and rearranged into the form

$$\mathbf{J} = \mathbf{J}_G + \mathbf{J}_{CS1} + \mathbf{J}_{CS2} \tag{5.35a}$$

and integrated with respect to time as follows. The first term can be rewritten with

$$\rho_{0\mathbf{r}_1\mathbf{r}_2}\rho_{0\mathbf{r}_2\mathbf{r}_3}\dot{\rho}'_{\mathbf{r}_3\mathbf{r}_1} = \partial_t(\rho_{0\mathbf{r}_1\mathbf{r}_2}\rho_{0\mathbf{r}_2\mathbf{r}_3}\rho'_{\mathbf{r}_3\mathbf{r}_1}) - \dot{\rho}_{0\mathbf{r}_1\mathbf{r}_2}\rho_{0\mathbf{r}_2\mathbf{r}_3}\rho'_{\mathbf{r}_3\mathbf{r}_1} - \rho_{0\mathbf{r}_1\mathbf{r}_2}\dot{\rho}_{0\mathbf{r}_2\mathbf{r}_3}\rho'_{\mathbf{r}_3\mathbf{r}_1}$$

and combined with the next two terms to give

$$\mathbf{J}_G = \frac{e}{\Omega}\partial_t \mathrm{Tr}[\rho_0, \mathbf{r}][\rho', \rho_0] \tag{5.35b}$$

and

$$\mathbf{J}_{CS1} = -3\frac{e}{\Omega}\mathrm{Tr}\rho'[\dot{\rho}_0, [\rho_0, \mathbf{r}]], \tag{5.35c}$$

while the final term in Eq. (5.34) (*i.e.*, the term involving ϕ_{123}) becomes

$$\mathbf{J}_{CS2} = -i\frac{e^2}{2\hbar\Omega}B^j\epsilon_{jab}\mathrm{Tr}[\rho_0, \mathbf{r}][\rho_0, r^a][r^b, \dot{\rho}_0] + \text{c.c.} \tag{5.35d}$$

upon rewriting

$$\begin{aligned}(\mathbf{r}_1 - 2\mathbf{r}_2 + \mathbf{r}_3)&(\mathbf{r}_1 \times \mathbf{r}_2 + \mathbf{r}_2 \times \mathbf{r}_3 + \mathbf{r}_3 \times \mathbf{r}_1) \\ &= (\mathbf{r}_1 - \mathbf{r}_2)[(\mathbf{r}_1 - \mathbf{r}_3) \times (\mathbf{r}_2 - \mathbf{r}_3)] + (\mathbf{r}_3 - \mathbf{r}_2)[(\mathbf{r}_1 - \mathbf{r}_2) \times (\mathbf{r}_1 - \mathbf{r}_3)].\end{aligned}$$

The total derivative term $\mathbf{J}_G$ [Eq. (5.35b)] can be written

$$J^i_G = \partial_t(\alpha_G)^i_j B^j \tag{5.36}$$

with α_G as given in Eq. (5.5b),

$$\begin{aligned}(\alpha_G)^i_j = \frac{e^2}{\hbar}\mathrm{Re}\sum_{\substack{n\,\mathrm{occ}\\ m\,\mathrm{unocc}}}\int_{\mathrm{BZ}}\frac{d^3k}{(2\pi)^3}&\frac{\langle u_n|\partial^i\mathcal{P}|u_m\rangle\langle u_m|\epsilon_{jab}\{\partial^a H, \partial^b\mathcal{P}\}|u_n\rangle}{E_n - E_m} \\ &+ 2e\,\mathrm{Im}\sum_{\substack{n\,\mathrm{occ}\\ m\,\mathrm{unocc}}}\int_{\mathrm{BZ}}\frac{d^3k}{(2\pi)^3}\frac{\langle u_n|\partial^i\mathcal{P}|u_m\rangle\langle u_m|\partial H'/\partial B^j|u_n\rangle}{E_n - E_m},\end{aligned} \tag{5.37}$$

where the dependence on $\mathbf{k}$ has been suppressed. This result follows immediately upon taking the trace in the basis of energy eigenstates. Matrix elements of $[\rho_0, r^i]$ appear as $\mathscr{r}^i_{\mathbf{k}} = \partial^i\mathcal{P}_k$,

from Eq. (5.27), and the cross-gap matrix elements of ρ' in are given in Eq. (5.28). Note that since $\mathbf{J}_G$ is a total time derivative, α_G is uniquely defined for a given Hamiltonian (this assumes the existence of a reference Hamiltonian with $\alpha_G = 0$, that is, the existence of a topologically trivial, time-reversal-invariant band insulator).

In $\mathbf{J}_{CS2}$ [Eq. (5.35d)], we can replace $r \to [[r, \rho_0], \rho_0]$ in the third commutator. This has the same cross-gap matrix elements as r; the interior matrix elements do not contribute to the trace because the other three factors, $\dot{\rho}_0$ and two components of $[\rho_0, r^i]$, have only cross-gap matrix elements. Then

$$\mathbf{J}_{CS2} = i\frac{e^2}{2\hbar\Omega} B^j \epsilon_{jab} \mathrm{Tr}[\rho_0, \mathbf{r}][\rho_0, r^a][[[\rho_0, r^b], \rho_0], \dot{\rho}_0] + \text{c.c.}$$

or

$$\mathbf{J}_{CS2} = -\frac{e^2}{2\hbar} B^j \epsilon_{jab} \mathrm{Tr} \boldsymbol{\nabla}_k \mathcal{P}_\mathbf{k} \partial^a \mathcal{P}_\mathbf{k} [[\partial^b \mathcal{P}_\mathbf{k}, \mathcal{P}_\mathbf{k}], \dot{\mathcal{P}}_\mathbf{k}] + \text{c.c.},$$

where an integral over $\mathbf{k}$ is suppressed for brevity and the trace is taken in the Hilbert space at $\mathbf{k}$. Dropping the subscripts $\mathbf{k}$ everywhere, this can be expanded and rearranged to give

$$\mathbf{J}_{CS2} = \frac{e^2}{2\hbar} B^j \epsilon_{jab} \mathrm{Tr}\mathcal{P}\{[\boldsymbol{\nabla}\mathcal{P}, \dot{\mathcal{P}}]\partial^a\mathcal{P}\partial^b\mathcal{P} + 2[\dot{\mathcal{P}}, \partial^b\mathcal{P}][\boldsymbol{\nabla}\mathcal{P}, \partial^a\mathcal{P}] + 3(\dot{\mathcal{P}}\partial^b\mathcal{P}\partial^a\mathcal{P}\boldsymbol{\nabla}\mathcal{P} + \partial^b\mathcal{P}\dot{\mathcal{P}}\boldsymbol{\nabla}\mathcal{P}\partial^a\mathcal{P})\}. \quad (5.35\text{d}')$$

In manipulating these strings of projection operators and their derivatives, it is very useful to realize that derivatives of projectors only have cross-gap matrix elements: $\mathcal{P}\partial^a\mathcal{P}\mathcal{P} = \mathcal{Q}\partial^a\mathcal{P}\mathcal{Q} = 0$, where $\mathcal{Q} = 1 - \mathcal{P}$ is the projector onto unoccupied bands. This means, for example, that $\mathcal{P}(\boldsymbol{\nabla}\mathcal{P})(\dot{\mathcal{P}}) = \mathcal{P}(\boldsymbol{\nabla}\mathcal{P})\mathcal{Q}(\dot{\mathcal{P}})\mathcal{P}$.

To $\mathbf{J}_{CS2}$ we must add $\mathbf{J}_{CS1}$ [Eq. (5.35c)],

$$\begin{aligned}\mathbf{J}_{CS1} &= \frac{3e^2}{2\hbar} B^j \epsilon_{jab} \mathrm{Tr}(\mathcal{P} - \mathcal{Q})[\dot{\mathcal{P}}, \boldsymbol{\nabla}\mathcal{P}]\partial^a\mathcal{P}\partial^b\mathcal{P} \\ &= \frac{3e^2}{2\hbar} B^j \epsilon_{jab} \mathrm{Tr}\mathcal{P}\{[\dot{\mathcal{P}}, \boldsymbol{\nabla}\mathcal{P}]\partial^a\mathcal{P}\partial^b\mathcal{P} - (\dot{\mathcal{P}}\partial^b\mathcal{P}\partial^a\mathcal{P}\boldsymbol{\nabla}\mathcal{P} + \partial^b\mathcal{P}\dot{\mathcal{P}}\boldsymbol{\nabla}\mathcal{P}\partial^a\mathcal{P})\}, \quad (5.35\text{c}')\end{aligned}$$

to get

$$\begin{aligned}\mathbf{J}_{CS} &= \mathbf{J}_{CS1} + \mathbf{J}_{CS2} \\ &= \frac{e^2}{\hbar} B^j \epsilon_{jab} \mathrm{Tr}\mathcal{P}\{[\dot{\mathcal{P}}, \boldsymbol{\nabla}\mathcal{P}]\partial^a\mathcal{P}\partial^b\mathcal{P} + [\dot{\mathcal{P}}, \partial^b\mathcal{P}][\boldsymbol{\nabla}\mathcal{P}, \partial^a\mathcal{P}]\}. \quad (5.38)\end{aligned}$$

By checking the different components explicitly one can see that this is

$$\mathbf{J}_{CS} = \mathbf{B}\frac{e^2}{\hbar} \mathrm{Tr}\mathcal{P}\{[\dot{\mathcal{P}}, \partial^x\mathcal{P}][\partial^y\mathcal{P}, \partial^z\mathcal{P}] + [\dot{\mathcal{P}}, \partial^y\mathcal{P}][\partial^z\mathcal{P}, \partial^x\mathcal{P}] + [\dot{\mathcal{P}}, \partial^z\mathcal{P}][\partial^x\mathcal{P}, \partial^y\mathcal{P}]\}, \quad (5.39)$$

so we get the "topological current"

$$\mathbf{J}_{CS} = -\mathbf{B}\frac{e^2}{\hbar} \int_{\mathrm{BZ}} \frac{d^3k}{(2\pi)^3} \mathrm{tr}(\mathcal{F}^{tx}\mathcal{F}^{yz} + \mathcal{F}^{ty}\mathcal{F}^{zx} + \mathcal{F}^{tz}\mathcal{F}^{xy}), \quad (5.40)$$

where the lower-case trace (tr) is only over the occupied bands, and the Brillouin-zone integral has been restored.

It remains only to show that $\mathbf{J}_{CS}$ is a total time derivative that integrates to $\alpha_{CS}\mathbf{B}$. Allowing the indices to run over t, x, y, z, in that order (so that $\epsilon_{txyz} = +1$),

$$\begin{aligned}\mathbf{J}_{CS} &= -\mathbf{B}\frac{e^2}{8\hbar}\int_{\mathrm{BZ}}\frac{d^3k}{(2\pi)^3}\epsilon_{abcd}\mathrm{tr}\mathcal{F}^{ab}\mathcal{F}^{cd} \\ &= -\mathbf{B}\frac{e^2}{2\hbar}\epsilon_{abcd}\int_{\mathrm{BZ}}\frac{d^3k}{(2\pi)^3}\partial^a\mathrm{tr}\left(\mathcal{A}^b\partial^c\mathcal{A}^d - i\frac{2}{3}\mathcal{A}^b\mathcal{A}^c\mathcal{A}^d\right). \end{aligned} \tag{5.41}$$

The derivatives with respect to k_x, k_y, k_z will vanish when integrated over the Brillouin zone assuming that $\mathcal{A}$ is defined smoothly and periodically over the zone, leaving just

$$\mathbf{J}_{CS} = -\mathbf{B}\frac{e^2}{2\hbar}\partial_t\int_{\mathrm{BZ}}\frac{d^3k}{(2\pi)^3}\epsilon_{abc}\mathrm{tr}\left(\mathcal{A}^a\partial^b\mathcal{A}^c - i\frac{2}{3}\mathcal{A}^a\mathcal{A}^b\mathcal{A}^c\right), \tag{5.42}$$

where the indices now only run over xyz, as originally. This obviously gives α_{CS} as in Eq. (5.5c), completing the proof. It must be reiterated that this integral is not always entirely trivial. In particular, if the adiabatic evolution brings the crystal back to its initial Hamiltonian in a nontrivial way, the Brillouin zone integral need not return to its initial value because $\mathcal{A}$ is not uniquely defined. In other words, $\int dt\mathbf{J}_{CS}$ can be multivalued as a function of the Hamiltonian deformation parameters. However, the change can only be such that θ changes by an integer multiple of 2π, as discussed in subsection 5.2.2.

5.4 Summary

The theoretical calculation of the magnetoelectric polarizability in insulators presents a difficulty similar to that known well from the theory of polarization; both quantities suffer an inherent ambiguity in the bulk. The magnetoelectric polarizability adds another level of difficulty because the vector potential is unbounded and breaks lattice translation symmetry. However, we have developed a formalism that allows us to deal directly with a uniform magnetic field. In the appendix, we further show that a long-wavelength regularization of the vector potential together with a suitable generalization of the polarization (to deal with the broken crystal symmetry) provides a (relatively) simple, though less rigorous, way to compute the response function. The final expression for the OMP rederives known results for particular model systems and topological insulators and completes the picture with additional terms that have a relatively straightforward and intuitive interpretation. We hope that these results and the method of their derivation will be valuable for future work on magnetoelectric effects and topological electronic phases.

Chapter 6

Conclusion

In the introduction it was claimed that the time-reversal-invariant topological insulators considered here, as well as the quantum Hall phases of matter, are generically characterized either by bulk properties (invariants) or edge properties (like the number of zero crossings of the edge dispersion). The foregoing chapters have focused almost exclusively on the bulk side of this correspondence in the case of topological insulators. Chapter 3 discussed a variant of the original two-dimensional topological invariant that is stable to disorder away from the clean transition, and that therefore serves to classify disordered (noninteracting, time-reversal-invariant) insulators. Chapters 4 and 5 discussed the magnetoelectric response properties of three-dimensional bulk insulators, which again give a topological invariant when time reversal symmetry $\mathcal{T}$ is not broken.

Much of the activity in the field of topological insulators to date, by contrast, has focused on the edge and surface states, not the bulk properties. This is certainly fair, since the edges carry low-energy excitations, which frequently dominate physical properties of materials in the experimental regimes of interest. Indeed, discussions of quantum Hall physics often maintain that the really "physical" characteristic is the chiral edge modes, since these modes carry the current in experimental setups.

However, in that case there are bulk measurements as well, like the low-frequency Faraday rotation, that detect the Hall conductance, so that having the bulk characterization is not irrelevant. Furthermore, as pointed out in Chapters 4 and 5, there is great interest in materials with bulk magnetoelectric properties, and so the results derived there may prove to be useful in the not-too-distant future. For example, in unpublished work, Sinisa Coh has already used them to compute the (small) orbital contributions to the magnetoelectric effect in the benchmark magnetoelectric Cr_2O_3. Hopefully a deeper understanding of topological insulators will lead to applications in this area.

Still, the fact that the work presented here excludes the boundaries by construction means that there should be more to the story. First, the presence of gapless surface states pose some difficulties for the derivations and numerics presented in the preceding chapters. Fundamentally, when there are gapless excitations none of the derivations in Chapters 3 or 5 are valid, since they utilize the adiabatic theorem, and hence its assumption of a spectral

gap. Of course, it is still possible to compute transport properties of gapless systems; the computation of the current through an insulator, as here, is actually a rather anomalous activity. The semiclassical formalism used in Chapter 4, for example, assumes that there is a gap but not that the chemical potential is in the gap; and there is a way to apply the Kubo formula to metals, although it is apparently a tricky business. So, it should be possible to compute the transport for a given metallic surface, but it needs to be said that it will be messier than the what has been explored here; the dissipationless transport considered here has a nice rigidity to it, which is lost in a metal.

As a specific example of the issues that can arise when the surfaces are taken into account, consider the following situation. As described in Chapter 4, the surface states of the topological insulator can be "gapped out" by applying a magnetic perturbation at the surface, and the sign of the surface Hall conductance then depends on the details of the perturbation. At an edge between two different signs of surface perturbation there is a change in the Hall conductance of $\pm e^2/h$, which means on very general grounds that the edge must carry a gapless chiral mode. Does this mean that the magnetoelectric polarizability of a topological insulator is generically not measurable, since one should expect to have such edge modes? Or should there be a some stability to a low density of such edges? It may be possible to treat this case by analogy with the integer quantum Hall effect, where disorder can lead to the presence of such modes in the bulk as well. However, the more general case, in which the surface Dirac modes are not gapped everywhere, may require a more novel treatment.

Another open question stems from the very different approaches taken to the topological invariant in Chapters 4 and 5 as compared to that of Chapter 3. The classification of topological insulators in Chapter 3 depended on the fact that $\mathcal{T}^2 = -1$ for spin-1/2 particles (electrons), and the three-dimensional invariant built from this two-dimensional invariant therefore seems to require this as well. This requirement is obscured in the formulation of the invariant as the k-space Chern-Simons integral, or the noncommutative volume introduced in the Interlude, but it can be seen (at least numerically) that the nontrivial value of the integral requires the sort of band degeneracy implied by $\mathcal{T}^2 = -1$ (*i.e.*, Kramers degeneracy). Even more strikingly, the crucial object in the derivation of Chapter 5 is the density matrix $\rho = \sum |\psi\rangle\langle\psi|$. If time reversal is represented by Θ, and $\Theta^2\psi = -\psi$, then $\Theta^2\rho(\Theta^{-1})^2 = \rho$ and the information about the spin-1/2 nature of the electron is apparently lost. On the other hand, the argument in Chapter 4 implying that $\mathcal{T}$-protected topological insulators exist even in the presence of electron-electron interactions does not depend at all on the spin of the electron (*i.e.*, Kramers degeneracy).

It may be that the only way to see the effect of strong spin-orbit coupling in general is to look at the surface modes. For example, this is how the 3D topological insulator is characterized in a recent treatment of disordered, noninteracting insulators [74]. Similarly, a bulk OMP of $\theta = \pi$ should be associated with something like gapless Dirac surface modes even in strongly interacting systems (recall, however, that the precise Dirac structure is not generic in the noninteracting, crystalline case, and so cannot be said to be truly fundamental).

It would be more satisfying to be able to trace the connection between the spin of the electron (*i.e.*, strong spin-orbit coupling) and the topological nature of the phase from bulk

considerations, even in the disordered and interacting cases. How does the requirement of degeneracy enter? In this regard, it is interesting to note that surface Hall conductances $\pm e^2/2h$ are consistent with the bulk topological insulator, even in the interacting case; that is, there seems to be a degeneracy lurking in plain sight. Will the bulk degeneracy depend on the topology of the material, as in the case of the fractional quantum Hall effect? That would be one way to reconcile the apparent discrepancy in the descriptions. In any case, the computation of the orbital magnetoelectric polarizability in the fully interacting case stands as an interesting and important open problem in this field.

Other questions involve the coupling of the electrons in topological and magnetoelectric materials to fields or perturbations other than the electric and magnetic fields, as well as to the electromagnetic field at higher orders. As an example, the old work on axion electrodynamics from particle physics points out that there will be higher order responses (F^4 rather than F^2, where F is the field tensor) suppressed by the electron "mass", *i.e.*, the band gap. At first glance, a strong magnetic field would seem to simply break the time-reversal invariance of the topological insulator and remove the interesting physics associated with that symmetry; these higher-order terms may indicate that a strong field may reveal other interesting, although probably not topological, effects.

This nonzero electric or magnetic field, at least when not too strong, can be seen as a perturbation away from the topological insulator, or more general OMP material. It is not immediately clear what sorts of perturbations will serve to enhance the magnetoelectric response of a material. The construction in Chapter 4, in which a coupling to an antiferromagnetic order parameter is added to the topological insulator, allows θ to be tuned continuously in that model, but in other cases such a coupling may have no effect at all on θ [55]. The non-Chern-Simons response α_G derived in Chapter 5 should generically be enhanced by perturbations that reduce the band gap, but since the Chern-Simons response depends crucially on the wave functions, it is not enough to look at the energy spectrum to understand the effects of varying the Hamiltonian.

On a similar note, it is interesting to ponder how the magnetoelectric response can actually be varied in the laboratory. After all, the derivation proceeds by assuming the possibility of such variation, at least in principle. Furthermore, it is constructed by analogy with the derivation of the modern theory of polarization [43], which is modeled in turn on the actual experimental protocol for determining changes in polarization. Finding a way to vary the Hamiltonian is like adding an extra dimension to the space in which the material lives, raising the possibility of effectively "looking in the bulk" without having to pass through the surface.

Finally, an area that has not been much explored to date is the question of the coupling between the electrons and other excitations in these materials. For starters, there is a theoretically intriguing, but hardly relevant experimentally, effect know from old work: a material with a nonzero value of θ should have a nontrivial coupling to the gravitational field, *i.e.*, the metric of spacetime [12].[1] While this sounds at first blush completely removed from

[1]While the effect due to coupling to the real gravitational field is highly unlikely to be measurable, it

the realm of condensed matter physics, it actually raises an interesting point, since the charge that couples to gravity is energy, and so the associated current is the energy, or heat, current in the material. It is know that the electronic properties of topological insulator thin films can affect the thermal currents in thin-film geometries [27]. What happens as a result of the coupling between the electrons and other excitations of the material, say, acoustic phonons or magnons? These are linearly dispersing bosons that couple to the electrons, like photons (and gravitons); it is therefore reasonable to expect that the unusual electronic structure that generates the magnetoelectric effect (and the "gravitational Hall effect") should also generate an interesting effect due to this coupling. More generally, orbital magnetoelectrics should be expected to serve as a background for exciting future developments in condensed matter physics.

should be noted that there are condensed matter models that have emergent gravitational fields [89], and it would be interesting to try to construct such a model that also had nonzero θ, in which case this effect would appear.

Bibliography

[1] B. L. Altshuler, A. G. Aronov, D. E. Khmel'nitskii, and A. I. Larkin. In I. M. Lifshits, editor, *Quantum Theory of Solids*. 1982.

[2] B. L. Altshuler, A. G. Aronov, A. I. Larkin, and D. E. Khmel'nitskii. *Zh. Eskp. Teor. Fiz.*, 81:768, 1981. [*Sov. Phys. JETP* **54**, 411 (1981)].

[3] J. E. Avron, L. Sadun, J. Segert, and B. Simon. Topological invariants in fermi systems with time-reversal invariance. *Phys. Rev. Lett.*, 61(12):1329–1332, Sep 1988.

[4] J. E. Avron, R. Seiler, and B. Simon. Homotopy and quantization in condensed matter physics. *Phys. Rev. Lett.*, 51(1):51–53, 1983.

[5] Joseph E. Avron and R. Seiler. Quantization of the Hall conductance for general, multiparticle Schr´odinger Hamiltonians. *Phys. Rev. Lett.*, 54(4):259–262, 1985.

[6] C. D. Batista, G. Ortiz, and A. A. Aligia. Ferrotoroidic moment as a quantum geometric phase. *Physical Review Letters*, 101(7):077203, 2008.

[7] G. Bergmann. Weak anti-localization–an experimental proof for the destructive interference of rotated spin 1/2. *Solid State Communications*, 42(11):815 – 817, 1982.

[8] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang. Quantum Spin Hall Effect and Topological Phase Transition in HgTe Quantum Wells. *Science*, 314:1757–, December 2006.

[9] M. V. Berry. Quantal phase-factors accompanying adiabatic changes. *Proc. Roy. Soc. A*, 392:45–57, 1984.

[10] E. Brown. Bloch electrons in a uniform magnetic field. *Phys. Rev.*, 133(4A):A1038–A1044, Feb 1964.

[11] Yu. A. Bychkov and É. I. Rashba. Properties of a 2d electron gas with lifted spectral degeneracy. *JETP Lett.*, 39(2):78–81, 1984.

[12] C. G. Callan Jr. and J. A. Harvey. Anomalies and fermion zero modes on strings and domain walls. *Nuclear Physics B*, 250(1-4):427 – 436, 1985.

[13] J. T. Chalker and P. D. Coddington. Percolation, quantum tunnelling and the integer Hall effect. *J. Phys. C: Solid State Phys.*, 21:2665–2679, 1988.

[14] Y. L. Chen, J. G. Analytis, J. H. Chu, Z. K. Liu, S. K. Mo, X. L. Qi, H. J. Zhang, D. H. Lu, X. Dai, Z. Fang, S. C. Zhang, I. R. Fisher, Z. Hussain, and Z.-X. Shen. Experimental Realization of a Three-Dimensional Topological Insulator, Bi_2Te_3. 325:178–181, 2009.

[15] Giulio D'Agostini. Bayesian inference in processing experimental data: Principles and basic applications. *Rept. Prog. Phys.*, 63:1383–1420, 2003.

[16] C. de Grandi and Anatoli Polkovnikov. Adiabatic perturbation theory: from landau-zener problem to quenching through a quantum critical point. in *Quantum Quenching, Annealing and Computation*, eds. A. Das, A. Chandra and B. K. Chakrabarti, Lect. Notes in Phys., vol. 802 (Springer, Heidelberg 2010).

[17] P. A. M. Dirac. Note on the interpretation of the density matrix in the many-electron problem. *Proc. Cam. Phil. Soc.*, 27:240–243, 1931. in *The collected works of P. A. M. Dirac, 1924-1928*, ed. R. H. Dalitz, Cambridge University Press (1995), p 497.

[18] M. I. D'yakonov and V. I. Perel. Possibility of orienting electron spins with current. *JETP Lett.*, 13:467, 1971.

[19] Claude Ederer and Nicola A. Spaldin. Towards a microscopic theory of toroidal moments in bulk periodic crystals. *Physical Review B*, 76(21):214404, 2007.

[20] Andrew M. Essin, Joel E. Moore, and David Vanderbilt. Magnetoelectric polarizability and axion electrodynamics in crystalline insulators. *Physical Review Letters*, 102(14):146805, 2009.

[21] M. Fiebig. Revival of the magnetoelectric effect. *J. Phys. D - Applied Phys.*, 38:R123–R152, 2005.

[22] L. Fu, C. L. Kane, and E. J. Mele. Topological insulators in three dimensions. *Phys. Rev. Lett.*, 98:106803, 2007.

[23] Liang Fu and C. L. Kane. Time reversal polarization and a Z_2 adiabatic spin pump. *Phys. Rev. B*, 74:195312, 2006.

[24] Liang Fu and C. L. Kane. Topological insulators with inversion symmetry. *Phys. Rev. B*, 76:045302, 2007.

[25] T. Fukui and Y. Hatsugai. Quantum spin Hall effect in three dimensional materials: Lattice computation of Z_2 topological invariants and its application to Bi and Sb. *J. Phys. Soc. Jpn.*, 76:053702, 2007.

[26] Takahiro Fukui and Yasuhiro Hatsugai. Quantum spin Hall effect in three dimensional materials: Lattice computation of Z_2 topological invariants and its application to Bi and Sb. *Phys. Rev. B*, 75:121403, 2007.

[27] Pouyan Ghaemi, Roger S. K. Mong, and Joel E. Moore. In-plane transport and enhanced thermoelectric performance in thin films of the topological insulators Bi_2Te_3 and Bi_2Se_3. arXiv:1002.1341, 2010.

[28] H.-M. Guo and M. Franz. Three-dimensional topological insulators on the pyrochlore lattice. *Physical Review Letters*, 103(20):206805, 2009.

[29] F. D. M. Haldane. Model for a quantum Hall effect without Landau levels: Condensed-matter realization of the "parity anomaly". *Physical Review Letters*, 61:2015–2018, October 1988.

[30] B. I. Halperin. Quantized Hall conductance, current-carrying edge states, and the existence of extended states in a two-dimensional disordered potential. *Phys. Rev. B*, 25:2185, 1982.

[31] M. Z. Hasan and C. L. Kane. Topological Insulators. unpublished; arXiv:1002.3895, 2010.

[32] F. W. Hehl, Y. N. Obukhov, J.-P. Rivera, and H. Schmid. Relativistic analysis of magnetoelectric crystals: Extracting a new 4-dimensional P odd and T odd pseudoscalar from Cr_2O_3 data. *Physics Letters A*, 372:1141–1146, February 2008.

[33] Shinobu Hikami. Localization, nonlinear σ model, and string theory. *Prog. Theor. Phys. Suppl.*, 107:213–227, 1992.

[34] Shinobu Hikami, Anatoly I. Larkin, and Yosuke Nagaoka. Spin-orbit interaction and magnetoresistance in the two dimensional random system. *Prog. Theor. Phys.*, 63:707–710, 1980.

[35] R. M. Hornreich and S. Shtrikman. Theory of gyrotropic birefringence. *Phys. Rev.*, 171(3):1065–1074, Jul 1968.

[36] Pavan Hosur, Shinsei Ryu, and Ashwin Vishwanath. Chiral topological insulators, superconductors, and other competing orders in three dimensions. *Phys. Rev. B*, 81:045120, 2010.

[37] D. Hsieh, D. Qian, L. Wray, Y. Xia, Y. S. Hor, R. J. Cava, and M. Z. Hasan. A topological Dirac insulator in a quantum spin Hall phase. *Nature*, 452:970, 2008.

[38] Daniel Huertas-Hernando, F. Guinea, and Arne Brataas. Spin-orbit coupling in curved graphene, fullerenes, nanotubes, and nanotube caps. *Phys. Rev. B*, 74(15):155426, Oct 2006.

[39] C. L. Kane and E. J. Mele. Quantum spin Hall effect in graphene. *Phys. Rev. Lett.*, 95:226801, 2005.

[40] C. L. Kane and E. J. Mele. Z_2 topological order and the quantum spin Hall effect. *Phys. Rev. Lett.*, 95:146802, 2005.

[41] Y. K. Kato, R. C. Myers, A. C. Gossard, and D. D. Awschalom. Observation of the spin Hall effect in semiconductors. *Science*, 306:1910, 2004.

[42] D. E. Khmelnitskii. Quantum hall effect and additional oscillations of conductivity in weak magnetic fields. *Physics Letters A*, 106(4):182 – 183, 1984.

[43] R. D. King-Smith and David Vanderbilt. Theory of polarization of crystalline solids. *Phys. Rev. B*, 47(3):1651–1654, Jan 1993.

[44] M. Koenig, S. Wiedmann, C. Bruene, A. Roth, H. Buhmann, L. W. Molenkamp, X.-L. Qi, and S.-C. Zhang. Quantum Spin Hall Insulator State in HgTe Quantum Wells. *Science*, 318:766, October 2007.

[45] Walter Kohn. Theory of the insulating state. *Phys. Rev.*, 133(1A):A171–A181, Jan 1964.

[46] R. B. Laughlin. Quantized Hall conductivity in two dimensions. *Phys. Rev. B*, 23:5632, 1981.

[47] R. B. Laughlin. Levitation of extended-state bands in a strong magnetic field. *Phys. Rev. Lett.*, 52(25):2304, Jun 1984.

[48] S.-S. Lee and S. Ryu. Many-Body Generalization of the Z_2 Topological Invariant for the Quantum Spin Hall Effect. *Physical Review Letters*, 100(18):186807–+, May 2008.

[49] I. B. Levinson. *Zh. Eskp. Teor. Fiz.*, 57:660, 1970. [*Sov. Phys. JETP* **30**, 362 (1970)].

[50] Rundong Li, Jing Wang, Xiaoliang Qi, and Shou-Cheng Zhang. 2009. arXiv:0908.1537.

[51] Andrei Malashevich, Ivo Souza, Sinisa Coh, and David Vanderbilt. Theory of orbital magnetoelectric response. unpublished; arXiv:1002.0300, 2010.

[52] Nicola Marzari and David Vanderbilt. Maximally localized generalized wannier functions for composite energy bands. *Phys. Rev. B*, 56(20):12847–12865, Nov 1997.

[53] R. McWeeny. Perturbation theory for the fock-dirac density matrix. *Phys. Rev.*, 126(3):1028–1034, May 1962.

[54] Hongki Min, J. E. Hill, N. A. Sinitsyn, B. R. Sahu, Leonard Kleinman, and A. H. MacDonald. Intrinsic and rashba spin-orbit interactions in graphene sheets. *Phys. Rev. B*, 74(16):165310, Oct 2006.

[55] Roger S. K. Mong, Andrew M. Essin, and Joel E. Moore. Antiferromagnetic topological insulators. arXiv:1004.1403, 2010.

[56] J. E. Moore and L. Balents. Topological invariants of time-reversal-invariant band structures. *Phys. Rev. B*, 75:121306(R), 2007.

[57] J. E. Moore, Y. Ran, and X.-G. Wen. Topological Surface States in Three-Dimensional Magnetic Insulators. *Physical Review Letters*, 101(18):186805–+, October 2008.

[58] Joel E. Moore. The birth of topological insulators. *Nature*, (464):194–198, 2010.

[59] S. Murakami, N. Nagaosa, and S.-C. Zhang. Dissipationless spin current at room temperature. *Science*, 301:1348, 2003.

[60] M. Nakahara. *Geometry, Topology and Physics*. Institute of Physics, 1998.

[61] Qian Niu and D. J. Thouless. Quantum Hall effect with realistic boundary conditions. *Phys. Rev. B*, 35(5):2188–2197, 1987.

[62] Qian Niu, D. J. Thouless, and Yong-Shi Wu. Quantized Hall conductance as a topological invariant. *Phys. Rev. B*, 31(6):3372–3377, 1985.

[63] Hideaki Obuse, Akira Furusaki, Shinsei Ryu, and Christopher Mudry. Two-dimensional spin-filtered chiral network modelof the $\mathbb{Z}_2$ quantum spin-Hall effect. *cond-mat/0702199*, 2007.

[64] Masaru Onoda, Yshai Avishai, and Naoto Nagaosa. Localization in a quantum spin Hall system. *Phys. Rev. Lett*, 98:076802, 2007.

[65] Gerardo Ortiz and Richard M. Martin. Macroscopic polarization as a geometric quantum phase: Many-body formulation. *Phys. Rev. B*, 49(20):14202–14210, May 1994.

[66] A. M. M. Pruisken. Universal singularities in the integral quantum hall effect. *Phys. Rev. Lett.*, 61(11):1297–1300, Sep 1988.

[67] Xiao-Liang Qi, Taylor L. Hughes, and Shou-Cheng Zhang. Topological field theory of time-reversal invariant insulators. *Physical Review B*, 78:195424, 2008.

[68] Xiao-Liang Qi and Shou-Cheng Zhang. The quantum spin hall effect and topological insulators. *Physics Today*, 63(1):33–38, 2010.

[69] E. I. Rashba, L. E. Zhukov, and A. L. Efros. Orthogonal localized wave functions of an electron in a magnetic field. *Phys. Rev. B*, 55:5306–5312, 1997.

[70] R. Resta. Theory of the electric polarization in crystals. *Ferroelectrics*, 136(1):51, 1992.

[71] Rahul Roy. $Z2$ classification of quantum spin hall systems: An approach using time-reversal invariance. *Phys. Rev. B*, 79(19):195321, May 2009.

[72] Rahul Roy. Topological phases and the quantum spin hall effect in three dimensions. *Phys. Rev. B*, 79(19):195322, May 2009.

[73] J. J. Sakurai. *Modern Quantum Mechanics*, chapter 4.4. Addison-Wesley, 1994.

[74] Andreas P. Schnyder, Shinsei Ryu, Akira Furusaki, and Andreas W. W. Ludwig. Classification of topological insulators and superconductors in three spatial dimensions. *Phys. Rev. B*, 78:195125, 2008.

[75] D. N. Sheng, Z. Y. Weng, L. Sheng, and F. D. M. Haldane. Quantum spin-hall effect and topologically invariant chern numbers. *Phys. Rev. Lett.*, 97(3):036808, Jul 2006.

[76] Jairo Sinova, Dimitrie Culcer, Q. Niu, N. A. Sinitsyn, T. Jungwirth, and A. H. MacDonald. Universal intrinsic spin hall effect. *Phys. Rev. Lett.*, 92(12):126603, Mar 2004.

[77] N. A. Spaldin and M. Fiebig. The renaissance of magnetoelectric multiferroics. *Science*, 309:391–392, 2005.

[78] T. Thonhauser, Davide Ceresoli, David Vanderbilt, and R. Resta. Orbital magnetization in periodic insulators. *Phys. Rev. Lett.*, 95(13):137205, Sep 2005.

[79] D. J. Thouless. Quantization of particle transport. *Phys. Rev. B*, 27(10):6083–6087, May 1983.

[80] D. J. Thouless. Wannier functions for magnetic sub-bands. *Journal of Physics C: Solid State Physics*, 17(12):L325, 1984.

[81] D. J. Thouless, M. Kohmoto, M. P. Nightingale, and M. den Nijs. Quantized Hall Conductance in a Two-Dimensional Periodic Potential. *Phys. Rev. Lett.*, 49:405, 1982.

[82] F. Wilczek. Two applications of axion electrodynamics. *Phys. Rev. Lett.*, 58:1799, 1987.

[83] Jacek C. Wojdeł and Jorge Íñiguez. Magnetoelectric response of multiferroic $bifeo3$ and related materials from first-principles calculations. *Phys. Rev. Lett.*, 103:267205, 2009.

[84] J. Wunderlich, B. Kaestner, J. Sinova, and T. Jungwirth. Experimental observation of the spin-hall effect in a two-dimensional spin-orbit coupled semiconductor system. *Phys. Rev. Lett.*, 94(4):047204, Feb 2005.

[85] Y. Xia, D. Qian, D. Hsieh, L. Wray, A. Pal, H. Lin, A. Bansil, D. Grauer, Y. S. Hor, R. J. Cava, and M. Z. Hasan. Observation of a large-gap topological-insulator class with a single Dirac cone on the surface. *Nature Physics*, 5:398–402, May 2009.

[86] Di Xiao, Ming-Che Chang, and Qian Niu. Berry phase effects on electronic properties. arXiv:0907.2021, 2009.

[87] Di Xiao, Junren Shi, Dennis P. Clougherty, and Qian Niu. Polarization and adiabatic pumping in inhomogeneous crystals. *Phys. Rev. Lett.*, 102:087602, 2009.

[88] Di Xiao, Junren Shi, and Qian Niu. Berry phase correction to electron density of states in solids. *Phys. Rev. Lett.*, 95(13):137204, Sep 2005.

[89] Cenke Xu. Gapless bosonic excitation without symmetry breaking: An algebraic spin liquid with soft gravitons. *Phys. Rev. B*, 74(22):224433, Dec 2006.

[90] Kun Yang and R. N. Bhatt. Floating of extended states and localization transistion in a weak magnetic field. *Phys. Rev. Lett.*, 76(8):1316–1319, 1996.

[91] W.-M. Yao, C. Amsler, D. Asner, R.M. Barnett, J. Beringer, P.R. Burchat, C.D. Carone, C. Caso, O. Dahl, G. D'Ambrosio, A. DeGouvea, M. Doser, S. Eidelman, J.L. Feng, T. Gherghetta, M. Goodman, C. Grab, D.E. Groom, A. Gurtu, K. Hagiwara, K.G. Hayes, J.J. Hernández-Rey, K. Hikasa, H. Jawahery, C. Kolda, Kwon Y., M.L. Mangano, A.V. Manohar, A. Masoni, R. Miquel, K. Mönig, H. Murayama, K. Nakamura, S. Navas, K.A. Olive, L. Pape, C. Patrignani, A. Piepke, G. Punzi, G. Raffelt, J.G. Smith, M. Tanabashi, J. Terning, N.A. Törnqvist, T.G. Trippe, P. Vogel, T. Watari, C.G. Wohl, R.L. Workman, P.A. Zyla, B. Armstrong, G. Harper, V.S. Lugovsky, P. Schaffner, M. Artuso, K.S. Babu, H.R. Band, E. Barberio, M. Battaglia, H. Bichsel, O. Biebel, P. Bloch, E. Blucher, R.N. Cahn, D. Casper, A. Cattai, A. Ceccucci, D. Chakraborty, R.S. Chivukula, G. Cowan, T. Damour, T. DeGrand, K. Desler, M.A. Dobbs, M. Drees, A. Edwards, D.A. Edwards, V.D. Elvira, J. Erler, V.V. Ezhela, W. Fetscher, B.D. Fields, B. Foster, D. Froidevaux, T.K. Gaisser, L. Garren, H.-J. Gerber, G. Gerbier, L. Gibbons, F.J. Gilman, G.F. Giudice, A.V. Gritsan, M. Grünewald, H.E. Haber, C. Hagmann, I. Hinchliffe, A. Höcker, P. Igo-Kemenes, J.D. Jackson, K.F. Johnson, D. Karlen, B. Kayser, D. Kirkby, S.R. Klein, K. Kleinknecht, I.G. Knowles, R.V. Kowalewski, P. Kreitz, B. Krusche, Yu.V. Kuyanov, O. Lahav, P. Langacker, A. Liddle, Z. Ligeti, T.M. Liss, L. Littenberg, L. Liu, K.S. Lugovsky, S.B. Lugovsky, T. Mannel, D.M. Manley, W.J. Marciano, A.D. Martin, D. Milstead, M. Narain, P. Nason, Y. Nir, J.A. Peacock, S.A. Prell, A. Quadt, S. Raby, B.N. Ratcliff, E.A. Razuvaev, B. Renk, P. Richardson, S. Roesler, G. Rolandi, M.T. Ronan, L.J. Rosenberg, C.T. Sachrajda, S. Sarkar, M. Schmitt, O. Schneider, D. Scott, T. Sjöstrand, G.F. Smoot, P. Sokolsky, S. Spanier, H. Spieler, A. Stahl, T. Stanev, R.E. Streitmatter, T. Sumiyoshi, N.P. Tkachenko, G.H. Trilling, G. Valencia, K. van Bibber, M.G. Vincter, D.R. Ward, B.R. Webber, J.D. Wells, M. Whalley, L. Wolfenstein, J. Womersley, C.L. Woody, A. Yamamoto, O.V. Zenin, J. Zhang, and R.-Y. Zhu. Review of Particle Physics. *J. Phys. G*, 33:1, 2006. §32.

[92] Yugui Yao, Fei Ye, Xiao-Liang Qi, Shou-Cheng Zhang, and Zhong Fang. Spin-orbit gap of graphene: First-principles calculations, Jan 2007.

[93] J. Zak. Magnetic translation group. *Phys. Rev.*, 134(6A):A1602–A1606, Jun 1964.

[94] Paolo Zanardi, Paolo Giorda, and Marco Cozzini. Information-theoretic differential geometry of quantum phase transitions. *Phys. Rev. Lett.*, 99(10):100603, Sep 2007.

Appendix A

Analogies between polarization and orbital magnetoelectric polarizability

	Polarization	Magnetoelectric polarizability
$d_{\min}$	1	3
EM coupling	$-\mathbf{P}\cdot\mathbf{E}$	$-\alpha\mathbf{E}\cdot\mathbf{B}$
Observable	$\mathbf{P}=\dfrac{\partial\langle H\rangle}{\partial E}$	$\alpha^i_j=-\dfrac{\partial\langle H\rangle}{\partial E_i\partial B^j}$
		$=\delta^i_j\dfrac{\theta}{2\pi}\dfrac{e^2}{h}$
Quantum	$\Delta\mathbf{P}=e\dfrac{\mathbf{R}}{\Omega}$	$\Delta\alpha=\dfrac{e^2}{h}$
Surface	$q=(\mathbf{P}_1-\mathbf{P}_2)\cdot\hat{\mathbf{n}}$	$\sigma_{xy}=(\alpha_1-\alpha_2)$
Chern-Simons form	$\mathcal{A}^i$	$\epsilon_{ijk}\left[\mathcal{A}^i\partial^j\mathcal{A}^k-i\dfrac{2}{3}\mathcal{A}^i\mathcal{A}^j\mathcal{A}^k\right]$

Table A.1: Comparison of Berry-phase theories of polarization and magnetoelectric polarizability.

Appendix B

Calculating the OMP using static polarization

As noted in the text, matrix elements of the operator r are ill-behaved in a basis of extended, Bloch-like states. That problem was solved by working with the density operator ρ, whose matrix elements are exponentially suppressed with distance. Another approach is to use a Wannier-like basis of localized states. In this appendix, we take this approach to present an alternative derivation of the OMP.

The Bloch functions $\psi_{n\mathbf{k}}(\mathbf{r})$ of the unperturbed crystal will evolve, under the application of a long-wavelength magnetic field $\mathbf{A} = \mathbf{A}_0 \sin \mathbf{q} \cdot \mathbf{r}$, into the exact energy eigenfunctions $\Psi_{n\mathbf{k}}(\mathbf{r})$. These no longer have a sharp crystal momentum $\mathbf{k}$, but may be expanded in a perturbation series in the unperturbed $\psi_{n\mathbf{k}}(\mathbf{r})$. Then the analogue to the standard Wannier function $w_{n\mathbf{R}}(\mathbf{r})$ for lattice vector $\mathbf{R}$ will be

$$\begin{aligned} W_{n\mathbf{R}}(\mathbf{r}) &= \frac{\Omega}{\sqrt{N}} \int_{\mathrm{BZ}} \frac{d^3k}{(2\pi)^3} \Psi_{n\mathbf{k}}(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{R}} \\ &= w_{n\mathbf{R}}(\mathbf{r}) + \delta w_{n\mathbf{R}}(\mathbf{r}), \end{aligned} \tag{B.1}$$

where Ω is the volume of the crystal and N is the number of unit cells. The Wannier orbitals centered at $\mathbf{R}$ become polarized when the magnetic field is applied, and this distortion gives a polarization density of

$$\delta\mathbf{P}(\mathbf{R}) = \frac{1}{\Omega} \sum_{n\,\mathrm{occ}} \langle w_{n\mathbf{R}} | e\mathbf{r} | \delta w_{n\mathbf{R}} \rangle + \text{c.c.} \tag{B.2}$$

Although it is not obvious that the bulk polarization appearing in Maxwell's equations is the same as the polarization of a set of Wannier orbitals, this expression leads to Eq. (5.5). To ensure that the Wannier orbitals are localized, we will have to suppose that each band has a vanishing Chern number,[80] so that the phase of $u_{n\mathbf{k}}$ can be chosen so that it is a periodic function of $\mathbf{k}$. In this case the unperturbed Wannier functions are localized, and (though there are usually subtleties in defining Wannier functions in a magnetic fields),[69] the regularization used here leads to localized orbitals. Presumably these arguments can be extended to

the case where the *total* Chern number for all occupied bands $C^{ij} = \sum_{n\,\mathrm{occ}} \int_{\mathrm{BZ}} d^3k \mathcal{F}^{ij}_{nn}(\mathbf{k})/2\pi$ vanishes.

Here we want to take a relatively direct approach to perturbation theory in the field, and write[88]

$$\begin{aligned} \Psi_{n\mathbf{k}} &= \psi_{n\mathbf{k}} + \delta\psi_{n\mathbf{k}} \\ \delta\psi_{n\mathbf{k}} &= \frac{eB}{2iq} \sum_l \left[\psi_{l\mathbf{k}+\mathbf{q}} \frac{\langle u_{l\mathbf{k}+\mathbf{q}} | v^x | u_{n\mathbf{k}} \rangle}{E_{n\mathbf{k}} - E_{l\mathbf{k}+\mathbf{q}} + i\epsilon} - (q \to -q) \right]. \end{aligned} \tag{B.3}$$

For definiteness we take $\mathbf{A} = -(B/q)\sin(qy)\hat{\mathbf{x}}$, and the velocity operator can be alternatively expressed as $v^x = \partial^x H_{\mathbf{k}}/\hbar$, with $H_{\mathbf{k}}$ the Bloch Hamiltonian of the unperturbed crystal.

Then the first-order correction to the dipole moment of the generalized Wannier functions will be

$$\delta P^i(\mathbf{R}) \;=\; e \sum_{n\,\mathrm{occ}} \int d\mathbf{r} \int_{\mathrm{BZ}} \frac{d^3k}{(2\pi)^3} \left(r^i e^{i\mathbf{k}\cdot(\mathbf{R}-\mathbf{r})} u^*_{n\mathbf{k}}(\mathbf{r}) \right) \int_{\mathrm{BZ}} \frac{d^3k'}{(2\pi)^3} \delta\psi_{n\mathbf{k}'}(\mathbf{r}) e^{-i\mathbf{k}'\cdot\mathbf{R}} + \text{c.c.} \tag{B.4}$$

The position integral must be taken over the whole crystal at this point. In the integral over k, r^i can be converted into a k derivative of the exponential, and then partial integration leaves a factor $-i\partial_{k_i} u^*$ (the boundary term vanishes because the Bloch function ψ is strictly periodic in k). Then

$$\delta P^i = -\frac{e^2 B}{2q} \sum_{\substack{n\,\mathrm{occ} \\ l}} \left[\frac{\langle \partial^i u_{n\mathbf{k}} | u_{l\mathbf{k}} \rangle \langle u_{l\mathbf{k}} | v^x | u_{n\mathbf{k}-\mathbf{q}} \rangle}{E_{n\mathbf{k}-\mathbf{q}} - E_{l\mathbf{k}} + i\epsilon} e^{i\mathbf{q}\cdot\mathbf{R}} - (q \to -q) \right] + \text{c.c.} \tag{B.5}$$

(From now on, we will omit the integral over $\mathbf{k}$ and the associated factor of $(2\pi)^3$.) Because of the variation of the magnetic field the magnetoelectric polarization $\delta P^i(\mathbf{R}) = \alpha^i_j B^j(\mathbf{R})$ should vary as $\cos \mathbf{q}\cdot\mathbf{R}$. The polarization seems to have both cosine and sine terms, but the coefficient of the latter is $-\mathbf{B}\sin(qy)C^{ix}/q$, and the vanishing of C is a prerequisite for using Wannier functions.

To lowest order in q and B, then, the magnetoelectric response is

$$\alpha^i_j = -\frac{e^2}{2} \epsilon_{jab} \partial_{q_b} \sum_{\substack{l \\ n\,\mathrm{occ}}} \frac{\langle \partial^i u_{n\mathbf{k}} | u_{l\mathbf{k}} \rangle \langle u_{l\mathbf{k}} | v^a | u_{n\mathbf{k}-\mathbf{q}} \rangle}{E_{n\mathbf{k}-\mathbf{q}} - E_{l\mathbf{k}} + i\epsilon} + \text{c.c.}, \tag{B.6}$$

where we have symmetrized over Landau gauges to make the expression nicer.

Switching to the shorthand $|n\rangle = |u_{n\mathbf{k}}\rangle$,

$$\alpha^i_j = \frac{e^2}{\hbar} \epsilon_{jab} \mathrm{Re} \sum_{\substack{n\,\mathrm{occ} \\ l}} \left[\frac{\langle \partial^i n | l \rangle \langle l | \partial^a H | \partial^b n \rangle}{E_n - E_l + i\epsilon} - \frac{\langle \partial^i n | l \rangle \langle l | \partial^a H | n \rangle \partial^b E_n}{(E_n - E_l + i\epsilon)^2} \right]. \tag{B.7}$$

Simplifying the second term of this expression makes use of the "Sternheimer equation"

$$(\partial^a H)|n\rangle = (E_n - H)|\partial^a n\rangle + (\partial^a E_n)|n\rangle \tag{B.8}$$

and the antisymmetry in the indices a and b to give

$$\alpha_j^i = \frac{e^2}{2\hbar}\epsilon_{jab}\sum_{\substack{n\,\mathrm{occ}\\ l}} \frac{\langle\partial^i n|l\rangle\langle l|\partial^a(H+E_n)|\partial^b n\rangle}{E_n - E_l + i\epsilon} + \text{c.c.} \tag{B.9}$$

Note the formal similarity to the expression for orbital magnetization,

$$M_j = \frac{1}{2}\mathrm{Im}\sum_{n\,\mathrm{occ}}\epsilon_{jab}\langle\partial^a n|(H+E_n)|\partial^b n\rangle = -\frac{1}{2}\mathrm{Im}\sum_{n\,\mathrm{occ}}\epsilon_{jab}\langle n|\partial^a(H+E_n)|\partial^b n\rangle, \tag{B.10}$$

in particular the appearance of the combination $H + E_n$.[78, 88]

To bring our compact expression into the form given in terms of α_G and α_{CS} in the main text, we need to break the sum over l into contributions from occupied and unoccupied states. Omitting the factor $(e^2/2\hbar)\epsilon_{jab}$ for the moment, the sum over the occupied states takes the form

$$\begin{aligned}\sum_{\substack{n,l\\ \mathrm{occ}}} \frac{\langle\partial^i n|l\rangle\langle l|\partial^a(H+E_n)|\partial^b n\rangle}{E_n - E_l + i\epsilon} + \text{c.c.} &= \sum_{\substack{n,l\\ \mathrm{occ}}}\langle\partial^i n|l\rangle\frac{\langle l|\partial^a(H-E_l)|\partial^b n\rangle + \langle\partial^b l|\partial^a(H-E_n)|n\rangle}{E_n - E_l + i\epsilon}\\ &= \sum_{\substack{n,n'\\ \mathrm{occ}}}\langle\partial^i n|n'\rangle\langle\partial^b n'|\partial^a n\rangle \end{aligned}\tag{B.11a}$$

using the antisymmetry in a and b and the Sternheimer equation again. Because the two sums are not symmetric when we take l in the unoccupied space, however, the terms do not cancel as nicely. Inserting a resolution of the identity, broken into two parts, gives:

$$\begin{aligned}&\sum_{\substack{n,n'\,\mathrm{occ}\\ m\,\mathrm{unocc}}} \frac{\langle\partial^i n|m\rangle\langle m|\partial^a H|n'\rangle\langle n'|\partial^b n\rangle}{E_n - E_m} + \text{c.c.}\\ &= \sum_{\substack{n,n'\,\mathrm{occ}\\ m\,\mathrm{unocc}}} \langle\partial^i n|m\rangle\langle m|\partial^a n'\rangle\langle n'|\partial^b n\rangle + \text{c.c.} &\text{(B.11b)}\\ &- \sum_{\substack{n,n'\,\mathrm{occ}\\ m\,\mathrm{unocc}}} \frac{\langle\partial^i n|m\rangle\langle m|\partial^a n'\rangle\langle n'|\partial^b H|n\rangle}{E_n - E_m} + \text{c.c.} &\text{(B.11c)}\\ &+ \sum_{\substack{n\,\mathrm{occ}\\ m\,\mathrm{unocc}}} \frac{\langle\partial^i n|m\rangle\langle m|\partial^a n\rangle\partial^b E_n}{E_n - E_m} + \text{c.c.}\end{aligned}$$

and

$$\sum_{\substack{n\,\mathrm{occ}\\ m,m'\,\mathrm{unocc}}} \frac{\langle\partial^i n|m\rangle\langle m|\partial^a H|m'\rangle\langle m'|\partial^b n\rangle}{E_n - E_m} + \text{c.c.} \tag{B.11d}$$
$$+ \sum_{\substack{n\,\mathrm{occ}\\ m\,\mathrm{unocc}}} \frac{\langle\partial^i n|m\rangle\langle m|\partial^b n\rangle\partial^a E_n}{E_n - E_m} + \text{c.c.}$$

The unnumbered pieces of these equations cancel by antisymmetry in a and b.

Defining $\mathcal{P}$ as the projector onto occupied bands as in the text, Eqs. (B.11c) and (B.11d) combine to give

$$(\alpha_G)^i_j = \frac{e^2}{2\hbar}\,\epsilon_{jab} \sum_{\substack{n\,\mathrm{occ}\\ m\,\mathrm{unocc}}} \frac{\langle n|\partial^i\mathcal{P}|m\rangle\langle m|\{\partial^a H, \partial^b\mathcal{P}\}|n\rangle}{E_n - E_m} + \text{c.c.}, \tag{B.12}$$

which is equivalent to Eq. (5.5b) upon identifying v^a with $\partial^a H$ and $\not{r}^i$ with $\partial^i\mathcal{P}$. This quantity has the crucial property that it is "gauge invariant," meaning that it can be written as a matrix trace, and hence does not change under a change of basis of the Hilbert space. Of course, this property is not evident here, where the formula makes explicit reference to energy eigenfunctions and their energies, but it follows from the expression in terms of a matrix given in Eq. (5.35b). The remainder, Eqs. (B.11a) and (B.11b), becomes

$$(\alpha_{\mathrm{CS}})^i_j = -\frac{e^2}{2\hbar}\delta^i_j\epsilon_{abc}\,\mathrm{tr}\left[\mathcal{A}^a\partial^b\mathcal{A}^c - \frac{2i}{3}\mathcal{A}^a\mathcal{A}^b\mathcal{A}^c\right], \tag{B.13}$$

which reproduces Eq. (5.5c).