An introduction to topological phases of electrons

Joel E. Moore, University of California, Berkeley, and
Lawrence Berkeley National Laboratory

OXFORD
UNIVERSITY PRESS

0.1 Introduction

These lectures seek to present a coherent picture of some key aspects of topological insulators and the quantum Hall effect. Rather than aiming for completeness or historical accuracy, the goal is to show that a few important ideas, such as the Berry phase and Chern and Chern-Simons differential forms, occur repeatedly and serve as links between superficially different areas of physics. Non-interacting topological phases, electrical polarization, and some transport phenomena in metals can all be understood in a unified framework as consequences of Abelian and non-Abelian Berry phases. The fractional quantum Hall effect is then discussed as an example of topological order, and we introduce its description by the (Abelian) Chern-Simons topological field theory.

Some effort has been made to avoid duplicating the material covered in other Les Houches lectures, both past and present. Readers seeking alternative approaches and a comprehensive list of references are encouraged to consult the many review articles on topological insulators [**?**, **?**, **?**] and the books by Bernevig [**?**] and Shen [**?**]. For the fractional quantum Hall effect, our treatment parallels closely the review articles of Wen [**?**, **?**], and the Les Houches notes of Girvin [**?**] provide an overview of the field and the theoretical background we will assume.

As part of our goal is to explain the topological invariants that underlie various topological phases, we start with a few examples of the two kinds of topology (homotopy and cohomology) that appear most frequently in condensed matter physics. No claims of rigor or completeness are made, and the book of Nakahara [**?**] is a good place to start learning more.

1
Basic concepts

1.1 Mathematical preliminaries

1.1.1 An intuitive example of global geometry and topology: Gauss-Bonnet

You may have heard a topologist described as "a mathematican who can't tell the difference between a donut and a coffee cup." As an example of the connections between geometry and topology, we start by discussing an integral that will help us classify two-dimensional compact manifolds (surfaces without boundaries) embedded smoothly in three dimensions. The integral we construct is "topologically invariant" in that if one such surface can be smoothly deformed into another, then the two will have the same value of the integral. The integral can't tell the difference between the surface of a coffee cup and that of a donut, but it can tell that the surface of a donut (a torus) is different from a sphere. Similar connections between global geometry and topology appear frequently in this course.

We start with a bit of local geometry. Given our 2D surface in 3D, we can choose coordinates at any point on the surface so that the $(x, y, z = 0)$ plane is tangent to the surface, which can locally be specified by a single function $z(x, y)$. We choose $(x = 0, y = 0)$ to be the given point, so $z(0, 0) = 0$. The tangency condition is

$$\frac{\partial z}{\partial x}\Big|_{(0,0)} = \frac{\partial z}{\partial y}\Big|_{(0,0)} = 0. \tag{1.1}$$

Hence we can approximate z locally from its second derivatives:

$$z(x, y) \approx \frac{1}{2} \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} \frac{\partial z}{\partial^2 x} & \frac{\partial z}{\partial x \partial y} \\ \frac{\partial z}{\partial y \partial x} & \frac{\partial z}{\partial^2 y} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{1.2}$$

The "Hessian matrix" that appears in the above is real and symmetric. It can be diagonalized and has two real eigenvalues λ_1, λ_2, corresponding to two orthogonal eigendirections in the (x, y) plane. The geometric interpretation of these eigenvalues is simple: their magnitude is an inverse radius of curvature, and their sign tells whether the surface is curving toward or away from the positive z direction in our coordinate system. To see why the first is true, suppose that we carried out the same process for a circle of radius r tangent to the x-axis at the origin. Parametrize the circle by an angle θ that is 0 at the origin and traces the circle counter-clockwise, i.e.,

$$x = r \sin \theta, \quad y = r(1 - \cos(\theta)). \tag{1.3}$$

Near the origin, we have

$$y = r(1 - \cos(\sin^{-1}(x/r)) = r - (1 - \frac{x^2}{2r^2}) = \frac{x^2}{2r}, \tag{1.4}$$

which corresponds to an eigenvalue $\lambda = 1/r$ of the matrix in Eq. 1.2.

Going back to the Hessian, its determinant (the product of its eigenvalues $\lambda_1\lambda_2$) is called the Gaussian curvature and has a remarkable geometric significance. First, consider a sphere of radius r, which at every point has $\lambda_1 = \lambda_2 = 1/r$. Then we can integrate the Gaussian curvature over the sphere's surface,

$$\int_{S^2} \lambda_1\lambda_2 \, dA = \frac{4\pi r^2}{r^2} = 4\pi. \tag{1.5}$$

Beyond simply being independent of radius, this integral actually gives the same value for any compact manifold that can be smoothly deformed to a sphere.

However, we can easily find a compact manifold with a different value for the integral. Consider the torus made by revolving the circle in Eq. 1.3, with $r = 1$, around the axis of symmetry $x = t, y = -1, z = 0$, with $-\infty < t < \infty$. To compute the Gaussian curvature at each point, we sketch the calculation of the eigenvalues of the Hessian as follows. One eigenvalue is around the smaller circle, with radius of curvature r: $\lambda_1 = 1/r = 1$. Then the second eigenvalue must correspond to the perpendicular direction, which has a radius of curvature that depends on the angle θ around the smaller circle (we keep $\theta = 0$ to indicate the point closest to the axis of symmetry). The distance from the axis of symmetry is $2 - \cos\theta$, so we might have guessed $\lambda_2 = (2 - \cos\theta)^{-1}$, but there is an additional factor of $\cos\theta$ that appears because of the difference in direction between the surface normal and this curvature. So our guess is that

$$\lambda_2 = -\frac{\cos\theta}{2 - \cos\theta} \tag{1.6}$$

As a check and to understand the sign, note that this predicts a radius of curvature 1 at the origin and other points closest to the symmetry axis, with a negative sign in the eigenvalue indicating that this curvature is in an opposite sense as that described by λ_1. At the top, the radius of curvature is 3 and in the same sense as that described by λ_1, and on the sides, λ_2 vanishes because the direction of curvature is orthogonal to the tangent vector.

Now we compute the curvature integral. With ϕ the angle around the symmetry axis, the curvature integral is

$$\int_{T^2} \lambda_1\lambda_2 \, dA = \int_0^{2\pi} d\theta \int_0^{2\pi} (2 - \cos\theta) \, d\phi \, \lambda_1\lambda_2 = \int_0^{2\pi} d\theta \int_0^{2\pi} d\phi \, (-\cos\theta) = 0. \tag{1.7}$$

Again this zero answer is generic to any surface that can be smoothly deformed to the torus. The general result (the Gauss-Bonnet formula) of which the above are examples is

$$\int_S \lambda_1\lambda_2 \, dA = 2\pi\chi = 2\pi(2 - g), \tag{1.8}$$

where χ is a topological invariant known as the Euler characteristic and g is the genus, essentially the number of "holes" in the surface. [1] For a compact manifold with boundaries, the Euler characteristic becomes $2 - 2g - b$, where b is the number of boundaries: one can check this by noting that by cutting a torus, one can produce two discs (by slicing a bagel) or alternately a cylinder with two boundaries (by slicing a bundt cake). We will not prove the Gauss-Bonnet formula but will encounter the Euler characteristic several times in these notes.

More generally, we will encounter several examples where a topological invariant is expressed as an integral over a local quantity with a geometric significance. We now turn to a simpler example in order to allow us to introduce some basic concepts of algebraic topology.

1.1.2 Invariant integrals along paths in two dimensions: exact forms

As our first example of a topological property, let's ask about making line integrals along paths (not path integrals in the physics sense, where the path itself is integrate over) that are nearly independent of the precise path: they will turn out to depend in some cases on topological properties (homotopy or cohomology). We will assume throughout these notes, unless otherwise specified, that all functions are smooth (i.e., $\mathbb{C}^\infty$, meaning derivatives of all orders exist).

First, suppose that we deal with paths on some open set U in the two-dimensional plane $\mathbb{R}^2$. (Open set: some neighborhood of each point in the set is also in the set.) We consider a smooth path $(u(t), v(t))$, where $0 \leq t \leq 1$ and the endpoints may be different. (To make these results more precise, we should provide for adding one path to another by requiring only piecewise smooth paths, and require that u and v be smooth in an open set including $t \in [0, 1]$. For additional rigor, see the first few chapters of W. Fulton, "Algebraic Topology: A First Course", Springer).

Now let $f(x, y) = (p(x, y), q(x, y))$ be a two-dimensional vector field that lets us compute line integrals of this path:

$$W = \int_0^1 dt\, p\frac{du}{dt} + q\frac{dv}{dt}dt, \tag{1.9}$$

where p and q are evaluated at $(x(t), y(t))$.

Mathematical note: in more fancy language, f is a differential form, a "1-form" to be precise. All that means is that f is something we can use to form integrals over paths that are linear and probe the tangent vector of the path. Another way to state this, with which you may be more familiar is that the tangent vector to a path, which we call a "vector", transforms naturally in an opposite way to the gradient of a function, which we call a "covector". To convince yourself that this is true, think about how both transform under a linear transformation on the underlying space. We will say a bit more about such forms in a moment.

[1] A good question is why we write the Euler characteristic as $2 - 2g$ rather than $1 - g$; one way to motivate this is by considering polygonal approximations to the surface. The discrete Euler characteristic $V - E + F$, where V, E, F count vertices, edges, and faces, is equal to χ. For example, the five Platonic solids all have $V - E + F = 2$.

Our first goal is to show that the following three statements are equivalent: (a) W depends only on the endpoints $(u(0), v(0))$ and $(u(1), v(1))$; (b) $W = 0$ for any closed path; (c) f is the gradient of a function g: $(p, q) = (\partial_x g, \partial_y g)$; The formal language used for (c) is that f is an *exact form*: $f = dg$ is the differential of a 0-form (a smooth function) g.

Note that (c) obviously implies (a) and (b), since then $W = g(u(1), v(1)) - g(u(0), v(0))$. To show that (b) implies (a), suppose (b) is true and (a) is not. Then there are two paths γ_1, γ_2 that have different integrals but the same endpoints. Form a new path γ so that, as t goes from 0 to $\frac{1}{2}$, γ_1 is traced, and then as t goes from $\frac{1}{2}$ to 1, γ_2 is traced opposite its original direction (now you can see why piecewise smooth paths are needed if one wants to be rigorous). Then this integral is nonzero, which contradicts (b).

It remains to show that (a) implies (c). Define $g(x, y)$ as equal to 0 at $(0, 0)$, or some other reference point in U if U does not include the origin. Everywhere else, set g equal to the W obtained by integrating over an arbitrary path from $(0, 0)$ to the final point, which by (a) is path-independent. (If U is not connected, then carry out this process on each connected component.) We will show that $\partial_x g = p$, and the same logic then implies $\partial_y g = q$. We need to compute

$$\partial_x g = \lim_{\Delta x \to 0} \frac{g(x + \Delta x, y) - g(x, y)}{\Delta x}. \tag{1.10}$$

We can obtain g by any path we like, so let's take an arbitrary path to define $g(x, y)$, then add a short horizontal segment to that path to define the path for $g(x + \Delta x, y)$. The value of the integral along this extra horizontal segment converges to $p(x, y)(\Delta x)$, as needed.

It turns out that the above case is simple because the plane we started with is "topologically trivial." Before proceeding to look at a nontrivial example, let us state one requirement on f that is satisfied whenever f is exact ($f = dg$). The fact that partial derivatives commute means that, with $f = dg = (p, q)$, $\partial_y p = \partial_x q$. We can come up with an elegant notation for this property by expanding our knowledge of differential forms.

Before, we obtained a 1-form f as the differential of a scalar g by defining

$$f = dg = \partial_x g \, dx + \partial_y g \, dy. \tag{1.11}$$

Note that we now include the differential elements dx, dy in the definition of f, and that 1-forms form a real vector space (spanned by dx, dy): we can add them and multiply them by scalars. To obtain a 2-form as the differential of a 1-form, we repeat the process: writing $f = f_i dx_i$ (with $x_1 = x, x_2 = y, f_1 = p, f_2 = q$)

$$df = \sum_j \frac{\partial f_i}{\partial x_j} \, dx_j \wedge dx_i. \tag{1.12}$$

where the $\wedge$ product between differential forms satisfies the rule $dx_i \wedge dx_j = -dx_j \wedge dx_i$, which implies that if any coordinate appears twice, then we get zero: $dx \wedge dx = 0$. For

some intuition about why this anticommutation property is important, note that in our 2D example,

$$df = (\partial_x f_y - \partial_y f_x) dx \wedge dy, \tag{1.13}$$

so that the function appearing in df is just the curl of the 2D vector field represented by f. So our statement about partial derivatives commuting is just the statement that if $f = dg$, then $df = 0$, or that the curl of a gradient is zero. We label any 1-form satisfying $df = 0$ a *closed form*. While every exact form is also closed, we will see that not every closed form is exact, with profound consequences.

1.1.3 Topologically invariant integrals along paths: closed forms

As an example of nontrivial topology, we would now like to come up with an example where integrals over paths are only path-independent in a limited "topological" sense: the integral is the same for any two paths that are *homotopic*, one of the fundamental concepts of topology (to be defined in a moment). Basically, two paths are homotopic if one can be smoothly deformed into another. Consider the vector field

$$f = (p, q) = \left(-\frac{y}{x^2 + y^2}, \frac{x}{x^2 + y^2} \right) = \frac{-y dx + x dy}{x^2 + y^2}, \tag{1.14}$$

where in the second step we have written it using our 1-form notation. This vector field is well-defined everywhere except the origin. This 1-form looks locally like the differential of $g = \tan^{-1}(y/x)$ (which just measures the angle in polar coordinates), but that function can only be defined smoothly on some open sets. For example, in a disc around the origin, the 2π ambiguity of the inverse tangent prevents defining g globally.

So if we have a path that lies entirely in a region where g can be defined, then the integral of this 1-form over the path will give the change in angle between the starting point and end point $g(u(1), v(1)) - g(u(0), v(0))$. What about other types of paths, for example, paths in $\mathbb{R}^2/\{0, 0\}$, the 2D plane with the origin omitted, that circle the origin and return to the starting point? We can still integrate using the 1-form f, even if it is not the gradient of a scalar function g, and will obtain the value $2\pi n$, where n is the "winding number": a signed integer that describes how many times the closed path $(u(t), v(t))$ circled the origin as t went from 0 to 1.

Now this winding number does not change as we make a small change in the closed path, as long as the path remains in $\mathbb{R}^2/\{0, 0\}$. What mathematical property of f guarantees this? Above we saw that any exact 1-form (the differential of a scalar function) is also closed. While f is not exact, we can see that it is closed:

$$df = \left(\partial_x \frac{x}{x^2 + y^2} \right) dx \wedge dy + \left(\partial_y \frac{-y}{x^2 + y^2} \right) dy \wedge dx = \frac{2 - 2}{x^2 + y^2} dx \wedge dy = 0. \tag{1.15}$$

In other words, $(-y, x)/(x^2+y^2)$ is curl-free ("irrotational"), while $(-y, x)$ has constant nonzero curl. Now suppose that we are given two paths γ_1 and γ_2 that differ by going in different ways around some small patch dA in which the 1-form remains defined. The difference in the integral of f over these two paths is then the integral of df over the enclosed surface by Stokes's theorem, which is zero if f is a closed form.

So we conclude that if f is a closed form, then the path integral is path-independent if we move the path through a region where f is always defined. For an exact form, the integral is completely path-independent. In the case of $\mathcal{R}/\{0,0\}$, the 1-form in Eq. 1.14 is locally but not completely path-independent. Both closed forms and exact forms are vector spaces (we can add and multiply by scalars), and typically infinite-dimensional, but their quotient as vector spaces is typically finite-dimensional. (The quotient of a vector space A by a vector space B is the vector space that identifies any two elements of A that differ only by an element of B). A basic object in "cohomology" is the first de Rham cohomology group (a vector space is by definition a group under addition),

$$H^1(M) = \frac{\text{closed 1-forms on } M}{\text{exact 1-forms on } M} = \frac{Z^1(M)}{B^1(M)}. \tag{1.16}$$

If you wonder why the prefix "co-" appears in "cohomology", there is a dual theory of linear combinations of curves, etc., called homology, in which the differential operator in de Rham cohomology is replaced by the boundary operator. However, while arguably more basic mathematically, homology seems to crop up less frequently in physics.

In this introductory discussion, we will focus on cohomology with real coefficients. The first and second Chern numbers defined later and applied to topological phases are actually elements of the even cohomology groups with *integer* coefficients $H^{2k}(M,\mathbb{Z})$. An even simpler object is the zeroth de Rham cohomology group. To understand this, realize that a closed 0-form is one whose gradient is zero, i.e., one that is constant on each connected component of U. There are no (-1)-forms and hence no exact 0-forms. So the zeroth group is just $\mathbb{R}^n$, where n is the number of connected components.

We can show that $H^1 = \mathbb{R}$ for the unit circle S^1 using the angle form f in Eq. 1.14, by showing that this form (more precisely, its equivalence class up to exact forms) provides a basis for H^1. Given some other form f', we use the unit circle path, parametrized by an angle θ going from zero to 2π, to define

$$c = \frac{\int_0^{2\pi} f'}{\int_0^{2\pi} f}. \tag{1.17}$$

Now $f' - cf$ integrates to zero. We can define a function g via

$$g(\theta) = \int_0^{\theta} (f' - cf). \tag{1.18}$$

Now g is well-defined and periodic because of how we defined c, and $f' = cf + dg$, which means that f' and cf are in the same equivalence class as dg is an exact form. We say that f' and f are cohomologous because they differ by an exact form. So cf, $c \in \mathbb{R}$, generates H^1, and $H^1(S^1)$ is isomorphic to $\mathbb{R}$. With a little more work, one can show that $\mathcal{R}/\{0,0\}$ also has $H^1 = \mathbb{R}$.

Actually we can connect the results of this section to the previous one: a general expression for the Euler characteristic is

$$\chi(M) = \sum_i (-1)^i \dim H^i(M) = \sum_i (-1)^i \dim \frac{Z^i(M)}{B_i(M)}. \tag{1.19}$$

The dimension of the ith cohomology group is called the ith Betti number (to be pedantic, the Betti numbers are defined for homology rather than cohomology, but we can use a duality relationship). There is a compact way to express the idea of cohomology and homology that will let us introduce some notation and terminology. If Ω_r is the vector space of r-forms, and C_r is the dual space of r-chains, then the action of the boundary operator and the differential is as follows:

$$\begin{aligned} &\longleftarrow C_r \xleftarrow[\partial_{r+1}]{} C_{r+1} \xleftarrow[\partial_{r+2}]{} C_{r+2} \longleftarrow \\ &\longrightarrow \Omega_r \xrightarrow[d_{r+1}]{} \Omega_{r+1} \xrightarrow[d_{r+2}]{} \Omega_{r+2} \longrightarrow . \end{aligned} \tag{1.20}$$

The rth cohomology group is the quotient $\ker d_{r+1}/\mathrm{im}\, d_r$, and the rth homology group is $\ker \partial_r/\mathrm{im}\, \partial_{r+1}$.

The duality relationship is provided by Stokes's theorem. Recall that this theorem relates the integral of a form over a boundary to the integral of the differential of the form over the interior. In terms of the linear operator (f, c) that evaluates the form f on the chain c, we have the compact expression

$$(f, \partial c) = (df, c). \tag{1.21}$$

Now we move on to a different type of topology that is perhaps more intuitive and will be useful for our first physics challenge: how to classify defects in ordered systems.

1.1.4 Homotopy

What if we did not want to deal with smooth functions and calculus? An even more basic type of topology is homotopy theory, which can be defined without reference to calculus, differential forms, etc. (although in physics the assumption of differentiability is usually applicable). Suppose that we are given a continuous map from $[0, 1]$ to a manifold M such that 0 and 1 get mapped to the same point; we can think of this as a closed curve on M. We say that two such curves γ_1, γ_2 are homotopic if there is a continuous function (a homotopy) f from $[0, 1] \times [0, 1]$ to M that satisfies

$$f(x, 0) = \gamma_1(x), \quad f(x, 1) = \gamma_2(x). \tag{1.22}$$

Intuitively, f describes how to smoothly distort γ_1 to γ_2. Now homotopy is an equivalence relation and hence defines equivalence classes: $[\gamma_1]$ is the set of all paths homotopic to γ_1. Furthermore, concatenation of paths (i.e., tracing one after the other) defines a natural group structure on these equivalence classes: the inverse of any path can be obtained by tracing it in the opposite direction. (To be precise, one should define homotopy with reference to a particular point where paths start and end; for a symmetric space where all points are basically equivalent, this is unnecessary.) We conclude that the equivalence classes of closed paths form a group $\pi_1(M)$, called the fundamental group or first homotopy group. Higher homotopy groups $\pi_n(M)$ are obtained by considering mappings from the n-sphere S^n to M in the same way.

The homotopy groups of a manifold are not independent of the cohomology groups: for example, if $\pi_1(M)$ is trivial, then so is the first de Rham group. The cohomology groups are always Abelian; in general, the first de Rham group *with integer coefficients*

is the Abelianization of π_1 (which need not be Abelian, although higher homotopy groups are). If you are interested in further details, the result of Hurewicz gives a relationship between higher cohomology and homotopy groups. The examples above of $\mathbb{R}^2/\{0,0\}$ and S^1 both have $\pi_1(M) = \mathbb{Z}$: there is an integer-valued winding number that we can use to classify paths, and this winding number can be computed by the angle form given above. So our two-dimensional examples already contains the two types of topology that occur most frequently in physics: de Rham cohomology and homotopy. We will return to homotopy in much more detail in a moment, when we explain how it can be used to classify topological defects such as vortices in broken-symmetry phases.

1.1.5 Application of homotopy to topological defects in symmetry-breaking phases

As a direct physical application of homotopy theory, consider the notion of a "vortex" in an ordered phase such as a superfluid. Such a configuration has a core where there is no order, but far away from the core the system is always locally in an ordered state. However, *which* ordered state the system is in varies smoothly as we move around the vortex core. For a 2D defect with a point core, such as a vortex of the 2D XY model, the vortex core is enclosed by a large real-space circle S^1, and as we move around this circle we have a map from S^1 to S^1, where the first circle is real space and the second circle reflects that the "order parameter manifold" of distinct ordered configurations of the XY model is also a circle.

The mathematical classification of topological defects has been carried out for a variety of systems. Vortex-like defects (defects that can be circled by a loop) are related to the group $\Pi_1(M)$, where M is the manifold of degenerate values of the order parameter once its magnitude has been set (for example, S^1 for XY and S^2 for Heisenberg, where S^d is the unit sphere in $d+1$ dimensions). $\pi_1(M)$ is known as the first homotopy group and is the group of equivalence classes of mappings from S^1 to M: for example, the mappings from S^1 to S^1 are characterized by an integer winding number $n \in \mathcal{Z}$, so $\pi_1(S^1) = \mathcal{Z}$, while $\pi_1(S^2) = 0$ (the group with one element) as any loop on the sphere is contractible to a point.

In other words, $\pi_1(M)$ gives the set of equivalence classes up to smooth deformations of closed paths on M. Multiplication of equivalence classes in the group is defined by concatenation of paths. The second homotopy group $\pi_2(M)$ classifies mappings from S^2 to M, and describes defects circled by a sphere, such as pointlike defects in 3D. For example, $\pi_2(S^2)$ is nonzero, and there are stable point defect configurations of Heisenberg spins (known descriptively as "hedgehogs") but not of XY spins. There can also be topological configurations that are not "defects" but not homotopic to the identity: the most famous example is the skyrmion configuration of a uniaxial ferromagnet in 2D, where all spins at infinity point in the same direction and the spin direction moves in the plane in such a way as to cover the sphere once. Shankar's monopoles and other defect-free configurations in 3D are related to the group π_3.

There is a considerable technology built up for the calculation of homotopy groups of general order parameter manifolds $M = G/H$, whose elements are cosets of the residual symmetry group H, i.e., any symmetries that survive in the ordered phase,

in the high-temperature symmetry group G. For example, for a uniaxially ordered Heisenberg ferromagnet, $G = SO(2)$ and $H = SO(3)$ so indeed $M = S^2$ as anticipated earlier. The advent of complicated ordered states in systems such as liquid crystals and spinor condensates stimulated the development of the techniques described in the Review of Modern Physics article by N. D. Mermin. [**?**]

1.2 Berry phases in quantum mechanics

We now turn to a beautiful geometric property of quantum mechanics that was understood relatively recently: the geometric or Berry phase. The connection to the Gauss-Bonnet theorem we mentioned earlier is as follows. Curvature is a property of Riemannian manifolds, which have a (real) inner product defined on the tangent space at each point. (The combination of a differentiable manifold and its tangent space at each point is the "tangent bundle", the simplest example of a vector bundle, an attachment of a vector space to each point of a manifold.) This inner product varies smoothly from point to point, which allows us to define a number of important concepts, including parallel transport and curvature.

Frequently in quantum mechanics we have, instead of a tangent space, a Hilbert space (including an Hermitian inner product) that varies smoothly from point to point in parameter space. Hence one can think of the Berry-phase objects we are about to define as really quite similar to curvature and related properties on Riemannian manifolds, except that the Berry phase does not come from the intrinsic geometry of the manifold of parameters but rather with how the attached Hilbert space evolves as parameters change.

An important result from undergraduate quantum mechanics is the "adiabatic approximation". Suppose that a system is prepared in a nondegenerate eigenstate of a time-dependent Hamiltonian H. For later reference, we will write H as a function of some parameters λ_i that depend on time: $H(t) = H(\lambda_1(t), \lambda_2(t), \ldots)$. If the eigenstate remains nondegenerate, then the adiabatic approximation is the result that if the Hamiltonian changes slowly in time (how slowly depends primarily on the energy gap between adjacent eigenstates), then there are no transitions between eigenstates.

This approximation, when correct, actually only gives part of the story: it describes the probability to remain in the eigenstate that evolved from the initial eigenstate, but there is actually nontrivial information in the *phase* of the final state as well. This result may seem quite surprising because the overall phase in quantum mechanics is in general independent of observable quantitites. However, the Berry phase from an adiabatic evolution is observable: for example, one system can be taken around a closed path in parameter space, while another system initially identical to the first can be taken around a different path, or the null path; an interference experiment on the final states will reveal the Berry phase. The first example of this type of geometric phase in physics was found more than fifty years ago by Pancharatnam in an optical example, but the classic Berry paper of 1984 was the first to explain the concept in its full generality.

Berry's result for a closed path is relatively simple to state, but some careful thought is required to understand and derive it. In moving a system adiabatically around a closed path in parameter space, the final wavefunction is in the same eigen-

state as the initial one (again, under the assumptions of the adiabatic approximation as stated above), but its phase has changed:

$$|\psi(t_f)\rangle = e^{-(i/\hbar)\oint_{t_i}^{t_f} E(t')\,dt'} e^{i\gamma} |\psi(t_i)\rangle. \tag{1.23}$$

Here $E(t')$ means the corresponding eigenvalue of the Hamiltonian at that time, and γ is the Berry phase, expressed as an integral over a path in *parameter* space with no time-dependence:

$$\gamma = i \int \langle \tilde{\psi}(\lambda_i) | \nabla_\lambda | \tilde{\psi}(\lambda_i) \rangle \cdot d\lambda. \tag{1.24}$$

Note that there are two different wavefunctions ψ and $\tilde{\psi}$ in the above formulas. $\psi(t)$ has a time argument and means the wavefunction of the system at that time. The "reference wavefunction" $\tilde{\psi}(\lambda_i)$ has a parameter argument and indicates the wavefunction we have chosen of the Hamiltonian for that value of the parameters, which we assume to be smoothly varying [2] A key assumption of the following derivation is that there is some smooth choice of the $\tilde{\psi}(\lambda_i)$ throughout a surface in parameter space with the loop as boundary.

For an open path, we need to describe the phase of the wavefunction relative to this reference set, so the expression becomes more complicated (for the closed path, we could simply compare the initial and final wavefunctions, without needing the reference set at these points). We will show that, assuming $\psi(t_i) = \tilde{\psi}(\lambda(t_i))$ so that the initial wavefunction is equal to the reference state at the corresponding value of parameters,

$$\langle \tilde{\psi}(\lambda_i(t)) | \psi(t) \rangle = e^{-(i/\hbar)\int_0^t E(t')\,dt'} e^{i\gamma} \equiv e^{i\theta(t)}, \tag{1.25}$$

i.e., the Berry phase appears as an extra contribution, beyond the expected contribution related to the energy, when comparing the actual time-dependent evolved state $\psi(t)$ to the reference state at the same point in parameter space $\lambda_i(t)$. We write $\theta(t)$ for the total phase including both energetic and Berry contributions. We can take the time derivative using the time-dependent Schrödinger equation

$$i\hbar \frac{\partial \psi}{\partial t} = H(t)\psi. \tag{1.26}$$

The first two quantities in (1.25) agree initially from our choice of the phase of $\psi(t_i)$. The time derivative of the leftmost is

$$\langle \tilde{\psi}(\lambda_i(t)) | \frac{-iE(t)}{\hbar} |\psi(t)\rangle + \frac{d\lambda_j}{dt} \langle \partial_{\lambda_j} \tilde{\psi}(\lambda_i(t)) | \psi(t) \rangle, \tag{1.27}$$

Since $e^{i\theta(t)} = \langle \psi(\lambda_i(t)) | \psi(t) \rangle$, this gives

$$i\partial_t \theta(t) = i \left[\frac{d}{dt} e^{i\theta(t)} \right] (-ie^{-i\theta(t)})$$

[2] A smooth choice of reference wavefunctions is always possible locally but not possible globally, as in the example of a spin-half particle moving in a Zeeman magnetic field.

$$= \left(\frac{-iE(t)}{\hbar} \langle \tilde{\psi}(\lambda_i(t))| + \frac{d\lambda_j}{dt} \langle \partial_{\lambda_j} |\tilde{\psi}(\lambda_i(t))| \right) |\psi(t)\rangle\langle\psi(t)|\tilde{\psi}(\lambda_i(t))\rangle, \quad (1.28)$$

and this is satisfied if we set (note that for E we do not need to distinguish between time and λ dependent)

$$\partial_t \theta(t) = -\frac{E(t)}{\hbar} - i\frac{d\lambda_j}{dt} \langle \partial_{\lambda_j} \tilde{\psi}(\lambda_i(t))|\tilde{\psi}(\lambda_i(t))\rangle, \quad (1.29)$$

which is our desired conclusion. We used the fact that ψ and $\tilde{\psi}$ differ only by a phase factor, since they describe the same non-degenerate state, to eliminate $|\psi\rangle\langle\psi|$.

The "Berry connection" or "Berry vector potential" $A_j = i\langle\psi(\lambda_i)|\partial_{\lambda_j}\psi(\lambda_i)\rangle$ is real, which follows from noting that $\partial_{\lambda_j}\langle\tilde{\psi}((\lambda_i)|\tilde{\psi}(\lambda_i)\rangle = 0$ by constancy of normalization. It is required for a nonzero Berry phase that H evolve in such a way that the wavefunction changes by more than just a phase, so that that the evolution of the wavefunction is more than just a simple phase factor, even though the actual rate of change in H drops out and only the path taken by H enters the Berry phase.

Now one can ask whether the Berry connection $\mathbf{A}$ is independent of how we chose the reference wavefunctions (in this case, the $U(1)$ degree of freedom in the wavefunction at each λ). While for an open path it clearly is not phase-independent, the Berry phase is phase-independent for a closed path, for exactly the same reasons as a closed line integral of $\mathbf{A}$ is gauge-independent in electrodynamics: we can integrate the "Berry flux" or "Berry curvature" $\epsilon_{ij}\partial_i A_j$ (which you can check is phase-independent, just like $F_{\mu\nu}$ in electrodynamics) on the surface bounded by the path. Alternately, we can note that a phase change changes A by the gradient of a scalar, so that on a closed loop, there is no change.

Independent of Berry's work and at roughly the same time, condensed matter physicists such as Thouless were realizing that Berry phases of wavefunctions on the Brillouin zone have the same mathematical structure of gauge fields in parameter space, even though there is no longer a notion of time evolution. The Berry vector potential $\mathbf{A}$ is a way to compare or "connect" the Hilbert spaces at neighboring points in parameter space. The gauge-invariant or nearly gauge-invariant quantities constructed from $\mathbf{A}$ and its derivatives control a variety of physical quantities. For the specific case of wavefunctions on the Brillouin zone, we will see that $\mathbf{A}$ is intimately related to the location of the wavefunctions within the unit cell in real space.

To get some geometric intuition for what the Berry phase means in general, we explain why the Berry connection A is called a connection, and the flux F is sometimes called a curvature. A connection is a way to compare vector spaces that are attached to different points of a manifold, forming a "vector bundle". In our case, there is a one-dimensional complex vector space attached at each point in parameter space, spanned by the local eigenstate. The inner product lets us compare vectors at the same point in parameter space, but the Berry connection appears when we try to compare two vectors from slightly different points.

An example we used above of a real vector bundle is the "tangent bundle" to a Riemannian manifold (say, a sphere), made up of tangent vectors at each point, which have a dot product corresponding to the inner product in quantum mechanics. The

connection in this case, which gives rise to "parallel transport" of tangent vectors, is related to the same curvature that we previously discussed with the Gauss-Bonnet theorem. Consider an airplane moving around the surface of the Earth and carrying a gyroscope that is fixed to lie in the tangent plane to the Earth's surface (i.e., free to rotate around the normal axis to the tangent plane). If the airplane follows a great circle, then it will appear to be going straight ahead to a passenger on board, and the gyroscope will not rotate relative to the plane's axis.

However, if the airplane follows a line of latitude other than the equator, or any other path that is not a "geodesic" (see a differential geometry text for details), it will feel constantly as though it is turning, and the gyroscope will appear to rotate relative to the airplane's direction. After going around a closed path in the airplane, the gyroscope may have rotated compared to a stationary gyroscope (the same physics that underlies Foucault's pendulum). As an exercise, you can work out that the total angle of rotation in circling a line of latitude is $2\pi \sin(\phi)$, where ϕ is the latitude. At the equator this gives no rotation, while at the north pole this gives a 2π rotation. This is a geometrical version of the same idea of holonomy (failure of a gyroscope to return to its initial direction) that underlies the Berry phase.

Note that a vector potential in a gauge theory and the associated Wilson loop are also examples of the concept of holonomy in a (now complex) vector bundle. The $U(1)$ Berry phase described above generalizes immediately to a non-Abelian $U(N)$ Berry phase when there are degenerate states or the energy differences between states are irrelevant, which has some important applications in condensed matter that were only recently discovered. Our primary mathematical objects in the following lectures will be properties of the wavefunctions on the Brillouin zone, which form a Hermitian bundle (a smoothly varying Hilbert space) on the d-dimensional torus.

One reason for introducing the idea of cohomology above was to give a sense for the mathematical structures hiding in the background of the simple calculations we do: to pick one example, the integral physicists do to calculate the Chern number, which determines the contribution of a filled 2D band to the quantum Hall effect, would be viewed by a mathematician as using the first Chern form to classify smooth complex line bundles on the Brillouin zone, and the group of line bundles under tensor products is isomorphic to the second cohomology class with integer coefficients. However, our hope is that the physical examples we discuss will be readily comprehensible even for readers not terribly excited about algebraic technology.

2

Topological phases I: Thouless phases arising from Berry phases

The integer quantum Hall effect has the remarkable property that, even at finite temperature in a disordered material, a transport quantity is quantized to remarkable precision: the transverse (a.k.a. Hall) conductivity is $\sigma_{xy} = ne^2/h$, where n is integral to 1 part in 10^9. This quantization results because the transport is determined by a topological invariant, as stated most clearly in work of Thouless. Consequently we use the term "Thouless phases" for phases where a response function is determined by a topological invariant.

In the cases we discuss, including the recently discovered "topological insulators" and quantum spin Hall effect, this topological invariant results from integration of an underlying Berry phase. It turns out that the Berry phase can be rather important even when it is not part of a topological invariant. In crystalline solids, the electrical polarization, the anomalous Hall effect, and the magnetoelectric polarizability all derive from Berry phases of the Bloch electron states, which are introduced in the following subsection. We will avoid the conventional textbook presentation of the IQHE in terms of Landau levels of a continuum electron. As we will use Landau levels when we discuss the fractional quantum Hall effect later, readers who are unfamiliar with the IQHE may wish to learn the standard treatment (see, e.g., Ref. [**?**]) and compare it to the approach using Bloch electrons below. The connection between the two can be made precise in the limit of small flux per unit cell, when a flat magnetic Bloch band becomes equivalent to a Landau level.

2.0.1 Bloch states

One of the cornerstones of the theory of crystalline solids is Bloch's theorem for electrons in a periodic potential. We will demonstrate this in the following form: given a potential invariant under a set of lattice vectors $\mathbf{R}$, $V(\mathbf{r}+\mathbf{R}) = V(\mathbf{r})$, the electronic eigenstates can be labeled by a "crystal momentum" $\mathbf{k}$ and written in the form

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}), \tag{2.1}$$

where the function u has the periodicity of the lattice. Note that the crystal momentum $\mathbf{k}$ is only defined up to addition of reciprocal lattice vectors, i.e., vectors whose dot product with any of the original lattice vectors is a multiple of 2π.

We give a quick proof of Bloch's theorem in one spatial dimension, then consider the Berry phase of the resulting wavefunctions. A standard fact from quantum mechanics tells us that, given two Hermitian operators that commute, we can find a basis

of simultaneous wavefunctions. In the problem at hand, we have a non-Hermitian operator (lattice translations by the lattice spacing a: $(T\psi)(x) = \psi(x+a)$) that commutes with the Hamiltonian. It turns out that only one of the two operators needs to be Hermitian for simultaneous eigenstates to exist, and therefore we can find wavefunctions that are energy eigenstates and satisfy

$$(T\psi)(x) = \lambda\psi(x). \tag{2.2}$$

Now if the magnitude of λ is not 1, repeated application of this formula will give a wavefunction that either blows up at spatial positive infinity or negative infinity. We would like to find wavefunctions that can extend throughout an infinite solid with bounded probability density, and hence require $|\lambda| = 1$. From that it follows that $\lambda = e^{i\theta}$, and we define $k = \theta/a$, where we need to specify an interval of width 2π to uniquely define θ, say $[-\pi, \pi)$. In other words, k is ambiguous by addition of a multiple of $2\pi/a$, as expected. So we have shown

$$\psi_k(x+a) = e^{ika}\psi_k(x). \tag{2.3}$$

The last step is to define $u_k(x) = \psi_k(x)e^{-ikx}$; then (2.3) shows that u_k is periodic with period a, and $\psi_k(x) = e^{ikx}u_k(x)$. [1]

While the energetics of Bloch wavefunctions underlies many properties of solids, there is also Berry-phase physics arising from the dependence of u_k on k that was understood only rather recently. Note that, even though this is one-dimensional, there is a nontrivial "closed loop" in the parameter k that can be defined because of the periodicity of the "Brillouin zone"' $k \in [-\pi/a, \pi/a)$:

$$\gamma = \oint_{-\pi/a}^{\pi/a} \langle u_k | i\partial_k | u_k \rangle dk. \tag{2.4}$$

How are we to interpret this Berry phase physically, and is it even gauge-invariant? We will derive it from scratch below, but an intuitive clue is provided if we make the replacement $i\partial_k$ by x, as would be appropriate if we consider the action on a plane wave. This suggests, correctly, that the Berry phase may have something to do with the spatial location of the electrons, but evaluating the position operator in a Bloch state gives an ill-defined answer; for this real-space approach to work, we would need to introduce localized "Wannier orbitals" in place of the extended Bloch states.

Another clue to what the phase γ might mean physically is provided by asking if it is gauge-invariant. Before, gauge-invariance resulted from assuming that the wavefunction could be continuously defined on the interior of the closed path. Here we have a closed path on a noncontractible manifold; the path in the integral winds around the Brillouin zone, which has the topology of the circle. What happens to the Berry phase if we introduce a phase change $\phi(k)$ in the wavefunctions, $|u_k\rangle \to e^{-i\phi(k)}|u_k\rangle$,

[1] Readers interested in more information and the three-dimensional case can consult the solid state text of Ashcroft and Mermin.

with $\phi(\pi/a) = \phi(-\pi/a) + 2\pi n, n \in \mathbb{Z}$? Under this transformation, the integral shifts as

$$\gamma \to \gamma + \oint_{-\pi/a}^{\pi/a} (\partial_k \phi)\, dk = \gamma + 2\pi n. \tag{2.5}$$

So redefinition of the wavefunctions shifts the Berry phase; we will see later that this corresponds to changing the polarization by a multiple of the "polarization quantum", which in one dimension is just the electron charge. (In higher dimensions, the polarization quantum is one electron charge per transverse unit cell.) Physically the ambiguity of polarization corresponds to the following idea: given a system with a certain bulk unit cell, there is an ambiguity in how that system is terminated and how much surface charge is at the boundary; adding an integer number of charges to one allowed termination gives another allowed termination (cf. Resta). The Berry phase is not gauge-invariant, but any fractional part it had in units of a *is* gauge-invariant. However, the above calculation suggests that, to obtain a gauge-invariant quantity, we need to consider a two-dimensional crystal rather than a one-dimensional one. Then integrating the Berry curvature, rather than the Berry connection, has to give a well-defined gauge-invariant quantity.

We will give a physical interpretation of γ in the next section as a one-dimensional polarization by relating changes in γ to electrical currents. (A generalization of this Berry phase is remarkably useful for the theory of polarization in real, three-dimensional materials.) In the next section we will understand how this one-dimensional example is related to the two-dimensional integer quantum Hall effect. Historically the understanding of Berry phases in the latter came first, in a paper by Thouless, Kohmoto, den Nijs, and Nightingale [**?**]. They found that, when a lattice is put in a commensurate magnetic field (one with rational flux per unit cell, in units of the flux quantum so that Bloch's theorem applies), each occupied band j contributes an integer

$$n_j = \frac{i}{2\pi} \int dk_x\, dk_y\, \left(\langle \partial_{k_x} u_j | \partial_{k_y} u_j \rangle - \langle \partial_{k_y} u_j | \partial_{k_x} u_j \rangle \right) \tag{2.6}$$

to the total Hall conductance:

$$\sigma_{xy} = \frac{e^2}{h} \sum_j n_j. \tag{2.7}$$

Now we derive this topological quantity (the "Chern number", expressed as an integral over the Berry flux, which is the curl of the Berry connection $A^j = i\langle u_j | \nabla_k u_j \rangle$) for the case of one-dimensional polarization, then explain its mathematical significance.

2.0.2 1D polarization and 2D IQHE

We start with the question of one-dimensional polarization mentioned earlier. More precisely, we attempt to compute the change in polarization by computing the integral of current through a bulk unit cell under an adiabatic change:

$$\Delta P = \int_0^1 d\lambda \frac{dP}{d\lambda} = \int_{t_0}^{t_1} dt\, \frac{dP}{d\lambda} \frac{d\lambda}{dt} = \int_{t_0}^{t_1} j(t)\, dt. \tag{2.8}$$

In writing this formula, we are assuming implicitly that there will be some definition of dP in terms of a parameter λ of the bulk Hamiltonian. Our treatment will follow that of Resta [**?**], but with a few more mathematical details in the derivation. (We write q for one-dimensional momentum and k_x, k_y for two-dimensional momenta in the following.) We will use Bloch's theorem in the following form: the periodic single-particle orbitals $u_n(q, r)$ are eigenstates of

$$H(q, \lambda) = \frac{1}{2m}(p + \hbar q)^2 + V^{(\lambda)}(r). \tag{2.9}$$

The current operator is

$$j(q) = ev(q) = \frac{ie}{\hbar}[H(q, \lambda), r] = \frac{e}{m}(p + \hbar q) = \frac{e}{\hbar}\partial_q H(q, \lambda). \tag{2.10}$$

The current at any fixed λ in the ground state is zero, but changing λ adiabatically in time drives a current that generates the change in polarization. To compute this current, we need to use the first correction to the adiabatic theorem (cf. the quantum mechanics book of Messiah). Following Thouless, we choose locally a gauge in which the Berry phase is zero (this can only be done locally and is only meaningful if we obtain a gauge-invariant answer for the instantaneous current), and write for the many-body wavefunction

$$|\psi(t)\rangle = \exp\left(-(i/\hbar)\int^t E_0(t')\,dt'\right)\left[|\psi_0(t)\rangle + i\hbar\sum_{j\neq 0}|\psi_j(t)\rangle(E_j - E_0)^{-1}\langle\psi_j(t)|\dot{\psi}_0(t)\rangle\right]. \tag{2.11}$$

Here $E_i(t)$ are the local eigenvalues and $|\psi_j(t)\rangle$ a local basis of reference states. The first term is just the adiabatic expression we derived before, but with the Berry phase eliminated with a phase rotation to ensure $\langle\psi_0(t)|\dot{\psi}_0(t)\rangle = 0$.

We want to use the above expression to write the expectation value of the current. The ground state must differ from the excited state by a single action of the (one-body) current operator, which promotes one valence electron (i.e., an electron in an occupied state) to a conduction electron. Using the one-particle states, we get

$$\frac{dP}{d\lambda} = 2\hbar e\,\mathrm{Im}\sum_{v,c}\int\frac{dq}{2\pi}\frac{\langle u_v(q)|v(q)|u_c(q)\rangle\langle u_c(q)|\partial_\lambda u_v(q)\rangle}{E_c(q) - E_v(q)}. \tag{2.12}$$

For example, we wrote

$$\langle\psi_j(t)|\dot{\psi}_0(t)\rangle = \sum_{v,c}\langle u_c|\partial_\lambda u_v\rangle\frac{d\lambda}{dt}. \tag{2.13}$$

This sum involves both valence and conduction states. For simplicity we assume a single valence state in the following. We can rewrite the sum simply in terms of the valence state using the first-order time-independent perturbation theory expression for the wavefunction change under a perturbation Hamiltonian $H' = dq\,\partial_q H$:

$$|\partial_q u_j(q)\rangle = \sum_{j\neq j'} |u_{j'}(q)\rangle \frac{\langle u_{j'}(q)|\partial_q H(q,\lambda)|u_j(q)\rangle}{E_j(q) - E_{j'}(q)}. \tag{2.14}$$

Using this and $v(q) = \frac{1}{\hbar}\partial_q H(q,\lambda)$ we obtain

$$\frac{dP}{d\lambda} = 2\hbar e \,\mathrm{Im} \sum_c \int \frac{dq}{2\pi} \frac{\langle u_v(q)|v(q)|u_c(q)\rangle\langle u_c(q)|\partial_\lambda u_v(q)\rangle}{E_c(q) - E_v(q)} = 2e \,\mathrm{Im} \int \frac{dq}{2\pi} \langle \partial_q u_v(q)|\partial_\lambda u_v(q)\rangle. \tag{2.15}$$

We can convert this to a change in polarization under a finite change in parameter λ:

$$\Delta P = 2e \,\mathrm{Im} \int_0^1 d\lambda \int \frac{dq}{2\pi} \langle \partial_q u_v(q)|\partial_\lambda u_v(q)\rangle. \tag{2.16}$$

The last expression is in two dimensions and involves the same type of integrand (a Berry flux) as in the 2D TKNN formula (2.6). However, in the polarization case there does not need to be any periodicity in the parameter λ. If this parameter is periodic, so that $\lambda = 0$ and $\lambda = 1$ describe the same system, then the total current run in a closed cycle that returns to the original Hamiltonian must be an integer number of charges, consistent with quantization of the TKNN integer in the IQHE.

If we define polarization via the Berry connection,

$$P = ie \int \frac{dq}{2\pi} \langle u_v(q)|\partial_q u_v(q)\rangle, \tag{2.17}$$

so that its derivative with respect to λ will give the result above with the Berry flux, we note that a change of gauge changes P by an integer multiple of the charge e. Only the fractional part of P is gauge-independent. The relationship between polarization in 1D, which has an integer ambiguity, and the IQHE in 2D, which has an integer quantization, is the simplest example of the relationship between Chern-Simons forms in odd dimension and Chern forms in even dimension. We now turn to the mathematical properties of these differential forms, which in the case above (and others to be discussed) came from the Berry phases of a band structure.

2.0.3 Interactions and disorder: the flux trick

One might worry whether the TKNN integer defined in equation (2.6) is specific to noninteracting electrons in perfect crystals. An elegant way to generalize the definition physically, while keeping the same mathematical structure, was developed by Niu, Thouless, and Wu [**?**]. This definition also makes somewhat clearer, together with our polarization calculation above, why this invariant should describe σ_{xy}. First, note that from the formula for the Bloch Hamiltonian in the polarization calculation above, we can reinterpret the crystal momentum q as a parameter describing a flux threaded through a unit cell of size a: the boundary conditions are periodic up to a phase $e^{iqa} = e^{ie\Phi/\hbar c}$. We will start by reinterpreting the noninteracting case in terms of such fluxes, then move to the interacting case.

The setup is loosely similar to the Laughlin argument for quantization in the IQHE. Consider adiabatically pumping a flux Φ_x though one circle of a toroidal system, in

the direction associated with the periodicity $x \to x + L_x, y \to y$. The change in this flux in time generates an electric field pointing in the $\hat{\mathbf{x}}$ direction. Treating this flux as a parameter of the crystal Hamiltonian, we compute the resulting change in $\hat{\mathbf{y}}$ polarization, which is related to the y current density:

$$\frac{dP_y}{dt} = j_y = \frac{dP_y}{d\Phi_x}\frac{d\Phi_x}{dt} = \frac{dP_y}{d\Phi_x}(cE_xL_x). \tag{2.18}$$

We are going to treat the polarization P_y as an integral over y flux but keep Φ_x as a parameter. Then (cf. Ortiz and Martin, 1994)

$$P_y(\Phi_x) = \frac{ie}{2\pi}\int d\Phi_y \, \langle u|\partial_{\Phi_y} u\rangle \tag{2.19}$$

and we see that polarization now has units of charge per length, as expected. In particular, the polarization quantum in the y direction is now one electronic charge per L_x. The last step to obtain the quantization is to assume that we are justified in averaging j_y over the flux:

$$\langle j_y\rangle = \langle\frac{dP_y}{d\Phi_x}\rangle(cE_xL_x) \to \frac{\Delta P_y}{\Delta\Phi_x}(cE_xL_x), \tag{2.20}$$

where Δ means the change over a single flux quantum: $\Delta\Phi_x = hc/e$. So the averaged current is determined by how many y polarization quanta change in the periodic adiabatic process of increasing the x flux by hc/e

$$\langle j_y\rangle = \frac{e}{hc}\frac{ne}{L_x}(cE_xL_x) = \frac{ne^2}{h}E_x. \tag{2.21}$$

The integer n follows from noting that computing $dP_y/d\Phi_x$ and then integrating $d\Phi_x$ gives just the expression for the TKNN integer (2.6), now in terms of fluxes.

2.0.4 TKNN integers, Chern numbers, and homotopy

In this section we will give several different ways to understand the TKNN integer or Chern number described above. First, a useful trick for many purposes is to define the Berry flux and first Chern number in a manifestly gauge-invariant way, using projection operators. For the case of a single non-degenerate band, define $P_j = |u_j\rangle\langle u_j|$ at each point of the Brillouin zone. This projection operator is clearly invariant under $U(1)$ transformations of u_j. The Chern number can be obtained as

$$n_j = \frac{i}{2\pi}\int \mathrm{Tr}\,[dP_j \wedge P_j\, dP_j]\,, \tag{2.22}$$

where $\wedge$ is the wedge product and $dP_j = \partial_{k_x}P_j\, dk_x + \partial_{k_y}P_j\, dk_y$ is a differential form where the coefficients are operators. (Note that the wedge product in the above formula acts only on dk_x and dk_y.) It is a straightforward exercise to verify that this reproduces the TKNN definition (2.6).

Then the generalization to degenerate bands, for example, is naturally studied by using the gauge- and basis-invariant projection operator $P_{ij} = |u_i\rangle\langle u_i| + |u_j\rangle\langle u_j|$ onto

the subspace spanned by $|u_i\rangle$ and $|u_j\rangle$: the index of this operator gives the total Chern number of bands i and j. In general, when two bands come together, only their total Chern number is defined. The total Chern number of all bands in a finite-dimensional band structure (i.e., a finite number of bands) is argued to be zero below. Often one is interested in the total Chern number of all occupied bands because this describes the integer quantum Hall effect through the TKNN formula; because of this zero sum rule, the total Chern number of all *unoccupied* bands must be equal and opposite.

In the remainder of this section, we use a powerful homotopy argument of Avron, Seiler, and Simon to show indirectly that there is one Chern number per band, but with a "zero sum rule" that all the Chern numbers add up to zero. We will not calculate the Chern number directly, but rather the homotopy groups of Bloch Hamiltonians. To get some intuition for the result, we first consider the example of a nondegenerate two-band band structure, then give the general result, which is an application of the "exact sequence of a fibration" mentioned in the Introduction.

The Bloch Hamiltonian for a two-band nondegenerate band structure can be written in terms of the Pauli matrices and the two-by-two identity as

$$H(k_x, k_y) = a_0(k_x, k_y)\mathbf{1} + a_1(k_x, k_y)\sigma_x + a_2(k_x, k_y)\sigma_y + a_3(k_x, k_y)\sigma_z. \tag{2.23}$$

The nondegeneracy constraint is that a_1, a_2, and a_3 are not all simultaneously zero. Now we first argue that a_0 is only a shift in the energy levels and has no topological significance, i.e., it can be smoothly taken to zero without a phase transition. Similarly we can deform the other a functions to describe a unit vector on $\mathbb{Z}_2$: just as the punctured plane $\mathbb{R}^2 - \{0, 0\}$ can be taken to the circle, we are taking punctured three-space to the two-sphere via

$$(a_1, a_2, a_3) \to \frac{(a_1, a_2, a_3)}{\sqrt{{a_1}^2 + {a_2}^2 + {a_3}^2}} \tag{2.24}$$

at each point in k-space.

Now we have a map from T^2 to S^2. We need to use one somewhat deep fact: under some assumptions, if $\pi_1(M) = 0$ for some target space M, then maps from the torus $T^2 \to M$ are contractible to maps from the sphere $S^2 \to M$. Intuitively this is because the images of the noncontractible circles of the torus, which make it different from the sphere, can be contracted on M. By this logic, the two-band nondegenerate band structure in two dimensions is characterized by a single integer, which can be viewed as the Chern number of the occupied band.

What is the Chern number, intuitively? For simplicity let's consider maps from S^2 to the non-degenerate two-band Hamiltonians described above. One picture is in terms of $\pi_2(S^2)$. A maybe more fundamental picture is that a nonzero Chern number is an "obstruction" to globally defining wavefunctions, in the following sense. F, the first Chern form, is a two-form. Let's consider a constant nonzero F, which for the case $S^2 \to S^2$ can be viewed as the field of a monopole located at the center of the target sphere. *Locally*, it is possible to find wavefunctions giving a vector potential A with $F = dA$, but not *globally*. (There has to be a "Dirac string" passing through the surface of the sphere somewhere.) In other words, states with nonzero Chern number

have Chern forms that are nontrivial elements of the second cohomology class: they are closed two-forms that are not globally exact.

The one subtle thing about this two-band model is that there is a nontrivial invariant in *three* spatial dimensions, since $\pi_3(S^2) = \mathbb{Z}$ (the "Hopf invariant"). In other words, even if the Chern numbers for the three two-dimensional planes in this three-dimensional structure are zero, there still can be an integer-valued invariant [2]. This map is familiar to physicists from the fact that the Pauli matrices can be used to map a normalized complex two-component spinor, i.e., an element of S^3, to a real unit vector, i.e., an element of S^2: $n^i = \mathbf{z}^\dagger \sigma^i \mathbf{z}$. This "Hopf map" is an example of a map that cannot be deformed to the trivial (constant) map. The Hopf invariant does not generalize to more than two bands, but what happens instead is quite remarkable.

Now we consider the case of a nondegenerate two-dimensional band structure with multiple bands, which we study using a method of Avron, Seiler, and Simon [**?**]. By the same argument as in the two-band case, we would like to understand π_1 and π_2 of the target space $H_{n\times n}$, nondegenerate $n \times n$ Hermitian matrices. As before, we will find that π_1 is zero so that maps from T^2 are equivalent to maps from S^2, but the latter will be quite nontrivial. We first diagonalize H at each point in k-space:

$$H(k) = U(k)D(k)U^{-1}(k). \tag{2.25}$$

Here $U(k)$ is unitary and $D(k)$ is real diagonal and nondegenerate. We can smoothly distort D everywhere in the Brillouin zone to a reference matrix with eigenvalues $1, 2, \ldots$ because of the nondegeneracy: if we plot the jth eigenvalue of D as a function of k_x and k_y, then this distortion corresponds to smoothing out ripples in this plot to obtain a constant plane.

The nontrivial topology is contained in $U(k)$. The key is to note that $U(k)$ in the above is ambiguous: right multiplication by any diagonal unitary matrix, an element of $DU(N)$, will give the same $H(k)$. So we need to understand the topology of $M = U(N)/DU(N) = SU(N)/SDU(N)$, where $SDU(N)$ means diagonal unitary matrices with determinant 1. We can compute π_2 of this quotient by using the exact sequence of a fibration and the following facts: $\pi_2(SU(N)) = \pi_1(SU(N)) = 0$ for $N \geq 2$. These imply that $\pi_2(M) \cong \pi_1(SDU(N)) = \mathbb{Z}^{n-1}$, i.e., $n-1$ copies of the integers. This follows from viewing $SDU(N)$ as N circles connected only by the requirement that the determinant be 1. Similarly we obtain $\pi_1(M) = 0$. We interpret these $n-1$ integers that arise in homotopy theory as just the Chern numbers of the bands, together with a constraint that the Chern numbers sum to zero.

2.0.5 Time-reversal invariance in Fermi systems

Now we jump to 2004-2005, when it was noted that imposing time-reversal symmetry in 2D electronic systems leads to new topological invariants. While nonzero Chern numbers cannot be realized with time-reversal invariance, the zero-Chern-number class gets subdivided into two pieces: "ordinary" insulators that do not in general have an edge state, and a "quantum spin Hall effect" or "topological insulator" where a bulk

[2]The nature of this fourth invariant changes when the Chern numbers are nonzero, as shown by Pontryagin in 1941: it becomes an element of a finite group rather than of the integers.

topological invariant forces an edge state. The topological invariant is not an integer here but rather a two-valued or $\mathbb{Z}_2$ invariant.

The idea that triggered this development started from considering two copies of the quantum Hall effect, one for spin-up electrons and one for spin-down, with opposite effective magnetic fields for the two spins. This combination, studied early on by Murakami, Nagaosa, Zhang [**?**], for example, is time-reversal invariant because acting with the time-reversal operator T changes both the magnetic field direction and the spin. Note that in a model such as this, S_z is a conserved quantum number even though $SU(2)$ (spin-rotation invariance) is clearly broken, as up and down spins behave differently. Heuristically, think of the spin-orbit coupling as arising from intra-atomic terms like $\mathbf{L} \cdot \mathbf{S}$, and consider specifically $L_z S_z$. For an electron of fixed spin, this coupling to the orbital motion described by L_z is just like the coupling in a constant magnetic field, since the orbital motion L_z generates a magnetic dipole moment. In the simplest case of a Chern number $+1$ state of up electrons and a Chern number -1 state of down electrons, the edge will have counterpropagating modes: e.g., up-spin moves clockwise along the edge and down-spin moves counterclockwise. This turns out not to be a bad caricature of the quantum spin Hall phase in a more realistic system: one can tell by symmetry arguments that it will have no quantum Hall effect (i.e., $\alpha_c = 0$ in $J_i = \alpha_c \epsilon_{ijk} E_j B_k$), it will have a spin Hall effect

$$J^i_j = \alpha_s \epsilon_{ijk} E_k, \tag{2.26}$$

where α_c and α_s are some numerical constants and J^i_j is a spin current (a current of angular momentum i in spatial direction j [3] The appearance of the electric field rather than the magnetic field in the quantum spin Hall equation results from the goal of having a potentially dissipationless current equation. If dissipation provides no "arrow of time", then both sides should transform in the same way under the time-reversal operation, which fixes the field on the right side to be E rather than B.

As an example of this "two copies of the IQHE" generated by spin-orbit coupling, consider the model of graphene introduced by Kane and Mele. [**?**] This is a tight-binding model for independent electrons on the honeycomb lattice (Fig. 2.1). 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:

$$H_0 = t \sum_{\langle ij \rangle \sigma} c^\dagger_{i\sigma} c_{j\sigma} + \lambda_v \sum_{i\sigma} \xi_i c^\dagger_{i\sigma} c_{i\sigma}. \tag{2.27}$$

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.

[3] There are some challenges that arise in trying to define a spin current in a realistic physical system, chiefly because spin is not a conserved quantity. Spin currents are certainly real and measurable in various situations, but the fundamental definition we give of the quantum spin Hall phase will actually be in terms of charge; "two-dimensional topological insulator" is a more precise description of the phase.

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

$$H_{SO} = i\lambda_{SO} \sum_{\langle\langle ij\rangle\rangle \sigma_1\sigma_2} \nu_{ij} c^\dagger_{i\sigma_1} s^z_{\sigma_1\sigma_2} c_{j\sigma_2}. \tag{2.28}$$

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. [**?**] Generic spin-orbit coupling in solids should not conserve any component of electron spin.

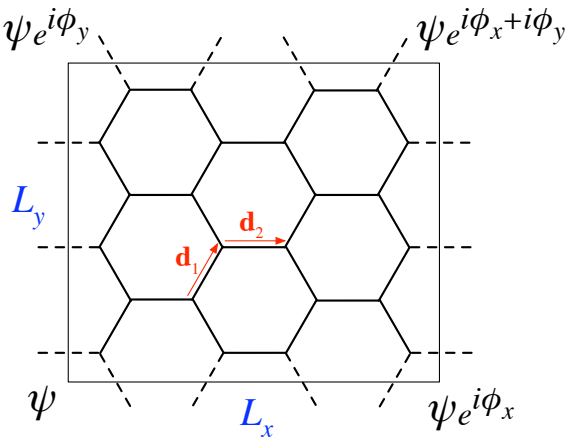

Fig. 2.1 (Color online) The honeycomb lattice on which the tight-binding Hamiltonian resides. For the two sites depicted, the factor ν_{ij} of equation (2.28) is $\nu_{ij} = -1$. The phases $\phi_{x,y}$ describe twisted boundary conditions that are used below to give a pumping definition of the $\mathbb{Z}_2$ invariant.

This model with S_z conservation is mathematically treatable using the Chern number above, as it just reduces to two copies of the IQHE. It is therefore not all that interesting in addition to not being very physical, because of the requirement of S_z conservation. In particular, the stability of the phase is dependent on a subtle property of spin-half particles (here we use the terms spin-half and Fermi interchangeably). The surprise is that the quantum spin Hall phase survives, with interesting modifications, once we allow more realistic spin-orbit coupling, as long as time-reversal symmetry remains unbroken.

The time-reversal operator T acts differently in Fermi and Bose systems, or more precisely in half-integer versus integer spin systems. Kramers showed long ago that the square of the time-reversal operator is connected to a 2π rotation, which implies that

$$T^2 = (-1)^{2S}, \tag{2.29}$$

where S is the total spin quantum number of a state: half-integer-spin systems pick up a minus sign under two time-reversal operations.

An immediate consequence of this is the existence of "Kramers pairs": every eigenstate of a time-reversal-invariant spin-half system is at least two-fold degenerate. We will argue this perturbatively, by showing that a time-reversal invariant perturbation H' cannot mix members of a Kramers pair (a state ψ and its time-reversal conjugate $\phi = T\psi$). To see this, note that

$$\langle T\psi|H'|\psi\rangle = \langle T\psi|H'|T^2\psi\rangle = -\langle T\psi|H'|\psi\rangle = 0, \tag{2.30}$$

where in the first step we have used the antiunitarity of T and the time-reversal symmetry of H', the second step the fact that $T^2 = -1$, and the last step is just to note that if $x = -x$, then $x = 0$.

Combining Kramers pairs with what is known about the edge state, we can say a bit about why a odd-even or $\mathbb{Z}_2$ invariant might be physical here. If there is only a single Kramers pair of edge states and we consider low-energy elastic scattering, then a right-moving excitation can only backscatter into its time-reversal conjugate, which is forbidden by the Kramers result above if the perturbation inducing scattering is time-reversal invariant. However, if we have two Kramers pairs of edge modes, then a right-mover can back-scatter to the left-mover that is *not* its time-reversal conjugate. This process will, in general, eliminate these two Kramers pairs from the low-energy theory.

Our general belief based on this argument is that a system with an even number of Kramers pairs will, under time-reversal-invariant backscattering, localize in pairs down to zero Kramers pairs, while a system with an odd number of Kramers pairs will wind up with a single stable Kramers pair. Additional support for this odd-even argument will be provided by our next approach. We would like, rather than just trying to understand whether the edge is stable, to predict from bulk properties whether the edge will have an even or odd number of Kramers pairs. Since deriving the bulk-edge correspondence directly is quite difficult, what we will show is that starting from the bulk T-invariant system, there are two topological classes. These correspond in the example above (of separated up- and down-spins) to paired IQHE states with even or odd Chern number for one spin. Then the known connection between Chern number and number of edge states is good evidence for the statements above about Kramers pairs of edge modes.

A direct Abelian Berry-phase approach for the 2D $\mathbb{Z}_2$ invariant is provided in the Appendix, along with an introduction to Wess-Zumino terms in 1+1-dimensional field theory and a physical interpretation of the invariant in terms of pumping cycles. The common aspect between these two is that in both cases the "physical" manifold (either the 2-sphere in the WZ case, or the 2-torus in the QSHE case) is extended in a certain way, with the proviso that the resulting physics must be independent of the precise nature of the extension. When we go to 3 dimensions in the following lecture, it turns out that there is a very nice 3D non-Abelian Berry-phase expression for the 3D $\mathbb{Z}_2$ invariant; while in practice it is certainly no easier to compute than the original expression based on applying the 2D invariant, it is much more elegant

mathematically so we will focus in that. Actually, for practical calculations, a very important simplification for the case of inversion symmetry (in both $d = 2$ and $d = 3$) was made by Fu and Kane: the topological invariant is determined by the product of eigenvalues of the inversion operator at the 2^d time-reversal symmetric points of the Brillouin zone. For further details we refer the reader to their 2007 PRB.

2.0.6 Experimental status of 2D insulating systems

This completes our discussion of one- and two-dimensional insulating systems. The two-dimensional topological insulator was observed by a transport measurement in $(Hg, Cd)Te$ quantum wells [**?**], following theoretical predictions [**?**]. A simplified description of this experiment is that it observed, in zero magnetic field, a two-terminal conductance $2e^2/h$, consistent with the expected conductance e^2/h for each edge if each edge has a single mode, with no spin degeneracy. More recent work has observed some of the predicted spin transport signatures as well, although as expected the amount of spin transported for a given applied voltage is not quantized, unlike the amount of charge.

In the next section, we start with the three-dimensional topological insulator and its remarkable surface and magnetoelectric properties. We then turn to metallic systems in order to understand another consequence of Berry phases of Bloch electrons.

2.0.7 3D band structure invariants and topological insulators

We will give a very quick introduction to the band structure invariants that allowed generalization of the previous discussion of topological insulators to three dimensions. However, most of our discussion of the three-dimensional topological insulator will be in terms of emergent properties that are difficult to perceive directly from the bulk band structure invariant. We start by asking to what extent the two-dimensional integer quantum Hall effect can be generalized to three dimensions. A generalization of the previous homotopy argument [**?**] can be used to show that there are three Chern numbers per band in three dimensions, associated with the xy, yz, and xz planes of the Brillouin zone. A more physical way to view this is that a three-dimensional integer quantum Hall system consists of a single Chern number and a reciprocal lattice vector that describes the "stacking" of integer quantum Hall layers. The edge of this three-dimensional IQHE is quite interesting: it can form a two-dimensional chiral metal, as the chiral modes from each IQHE combine and point in the same direction.

Consider the Brillouin zone of a three-dimensional time-reversal-invariant material. Our approach will be to build on our understanding of the two-dimensional case: concentrating on a single band pair, there is a $\mathbb{Z}_2$ topological invariant defined in the two-dimensional problem with time-reversal invariance. Taking the Brillouin zone to be a torus, there are two inequivalent xy planes that are distinguished from others by the way time-reversal acts: the $k_z = 0$ and $k_z = \pm\pi/a$ planes are taken to themselves by time-reversal (note that $\pm\pi/a$ are equivalent because of the periodic boundary conditions). These special planes are essentially copies of the two-dimensional problem, and we can label them by $\mathbb{Z}_2$ invariants $z_0 = \pm 1$, $z_{\pm 1} = \pm 1$, where $+1$ denotes "even Chern parity" or ordinary 2D insulator and -1 denotes "odd Chern parity" or

topological 2D insulator. Other xy planes are not constrained by time-reversal and hence do not have to have a $\mathbb{Z}_2$ invariant.

The most interesting 3D topological insulator phase (the "strong topological insulator") results when the z_0 and $z_{\pm 1}$ planes are in different 2D classes. This can occur if, moving in the z direction between these two planes, one has a series of 2D problems that interpolate between ordinary and topological insulators by breaking time-reversal. We will concentrate on this type of 3D topological insulator here. Another way to make a 3D topological insulator is to stack 2D topological insulators, but considering the edge of such a system shows that it will not be very stable: since two "odd" edges combine to make an "even" edge, which is unstable in the presence of T-invariant backscattering, we call such a stacked system a "weak topological insulator".

Above we found two xy planes with two-dimensional $\mathbb{Z}_2$ invariants. By the same logic, we could identify four other such invariants x_0, $x_{\pm 1}$, y_0, $y_{\pm 1}$. However, not all six of these invariants are independent: some geometry (exercise) shows that there are two relations, reducing the number of independent invariants to four:

$$x_0 x_{\pm 1} = y_0 y_{\pm 1} = z_0 z_{\pm 1}. \tag{2.31}$$

(Sketch of geometry: to establish the first equality above, consider evaluating the Fu-Kane 2D formula on the four EBZs described by the four invariants x_0, x_{+1}, y_0, y_{+1}. These define a torus, on whose interior the Chern two-form F is well-defined. Arranging the four invariants so that all have the same orientation, the A terms drop out, and the F integral vanishes as the torus can be shrunk to a loop. In other words, for some gauge choice the difference $x_0 - x_{+1}$ is equal to $y_0 - y_{+1}$.) We can take these four invariants in three dimensions as $(x_0, y_0, z_0, x_0 x_{\pm 1})$, where the first three describe layered "weak" topological insulators, and the last describes the Alternately, the "axion electrodynamics" field theory in the next subsection can be viewed as suggesting that there should be only one genuinely three-dimensional $\mathbb{Z}_2$ invariant.

For example, the strong topological insulator cannot be realized in any model with S_z conservation, while, as explained earlier, a useful example of the 2D topological insulator (a.k.a. "quantum spin Hall effect") can be obtained from combining IQHE phases of up and down electrons. The impossibility of making an STI with S_z conservation follows from noting that all planes normal to z have the same Chern number, as Chern number is a topological invariant whether or not the plane is preserved by time-reversal. In particular, the $k_z = 0$ and $k_z = \pm\pi/a$ phases have the same Chern number for up electrons, say, which means that these two planes are either both 2D ordinary or 2D topological insulators.

While the above argument is rigorous, it doesn't give much insight into what sort of gapless surface states we should expect at the surface of a strong topological insulator. The answer can be obtained by other means (some properties can be found via the field-theory approach given in the next section): the spin-resolved surface Fermi surface encloses an odd number of Dirac points. In the simplest case of a single Dirac point, believed to be realized in Bi_2Se_3, the surface state can be pictured as "one-quarter of graphene." Graphene, a single layer of carbon atoms that form a honeycomb lattice, has two Dirac points and two spin states at each k; spin-orbit coupling is quite weak since carbon is a relatively light element. The surface state of a three-dimensional

topological insulator can have a single Dirac point and a single spin state at each k. As in the edge of the 2D topological insulator, time-reversal invariance implies that the spin state at k must be the T conjugate of the spin state at $-k$.

2.0.8 Axion electrodynamics, second Chern number, and magnetoelectric polarizability

The three-dimensional topological insulator turns out to be connected to a basic electromagnetic property of solids. We know that in an insulating solid, Maxwell's equations can be modified because the dielectric constant ϵ and magnetic permeability μ need not take their vacuum values. Another effect is that solids can generate the electromagnetic term

$$\Delta\mathcal{L}_{EM} = \frac{\theta e^2}{2\pi h}\mathbf{E}\cdot\mathbf{B} = \frac{\theta e^2}{16\pi h}\epsilon^{\alpha\beta\gamma\delta}F_{\alpha\beta}F_{\gamma\delta}. \tag{2.32}$$

This term describes a magnetoelectric polarizability: an applied electrical field generates a magnetic dipole, and vice versa. An essential feature of the above "axion electrodynamics" theory (cf. Wilczek, 1987 [**?**]) is that, when the axion field $\theta(\boldsymbol{x},t)$ is constant, it plays no role in electrodynamics; this follows because θ couples to a total derivative, $\epsilon^{\alpha\beta\gamma\delta}F_{\alpha\beta}F_{\gamma\delta} = 2\epsilon^{\alpha\beta\gamma\delta}\partial_\alpha(A_\beta F_{\gamma\delta})$ (here we used that F is closed, i.e., $dF = 0$), 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 $\theta(\boldsymbol{x})$ appear.

A bit of work shows that, at a surface where θ changes, there is a surface quantum Hall layer of magnitude

$$\sigma_{xy} = \frac{e^2(\Delta\theta)}{2\pi h}. \tag{2.33}$$

(This can be obtained by moving the derivative from one of the A fields to act on θ, leading to a Chern-Simons term for the EM field at the surface. The connection between Chern-Simons terms and the quantum Hall effect will be a major subject of the last part of this course.) The magnetoelectric polarizability described above can be obtained from these layers: for example, an applied electric field generates circulating surface currents, which in turn generate a magnetic dipole moment. In a sense, σ_{xy} is what accumulates at surfaces because of the magnetoelectric polarizability, in the same way as charge is what accumulates at surfaces because of ordinary polarization.

We are jumping ahead a bit in writing θ as an angle: we will see that, like polarization, θ is only well defined as a bulk property modulo 2π (for an alternate picture on why θ is periodic that is more appropriate for electroweak symmetry breaking, see Ref. [**?**]). The integer multiple of 2π is only specified once we specify a particular way to make the boundary. How does this connect to the 3D topological insulator? At first glance, $\theta = 0$ in any time-reversal-invariant system, since $\theta \to -\theta$ under time-reversal. However, since θ is periodic, $\theta = \pi$ also works, as $-\theta$ and θ are equivalent because of the periodicity, and is inequivalent to $\theta = 0$.

Here we will not give a microscopic derivation of how θ can be obtained, for a band structure of noninteracting electrons, as an integral of the Chern-Simons form:

$$\theta = \frac{1}{2\pi} \int_{\mathrm{BZ}} d^3k \; \epsilon_{ijk} \, \mathrm{Tr}[\mathcal{A}_i \partial_j \mathcal{A}_k - i\frac{2}{3} \mathcal{A}_i \mathcal{A}_j \mathcal{A}_k], \tag{2.34}$$

which can be done by imitating our previous derivation of the polarization formula [**?**,**?**, **?**,**?**]. Instead we will focus on understanding the physical and mathematical meaning of the Chern-Simons form that constitutes the integrand, chiefly by discussing analogies with our previous treatment of polarization in one dimension and the IQHE in two dimensions. These analogies are summarized in Table I.

Throughout this section,

$$\mathcal{F}_{ij} = \partial_i \mathcal{A}_j - \partial_j \mathcal{A}_i - i[\mathcal{A}_i, \mathcal{A}_j] \tag{2.35}$$

is the (generally non-Abelian) Berry curvature tensor ($\mathcal{A}_\lambda = i\langle u | \partial_\lambda | u \rangle$), and the trace and commutator refer to band indices. We will understand the Chern-Simons form $K = \mathrm{Tr}[\mathcal{A}_i \partial_j \mathcal{A}_k - i\frac{2}{3} \mathcal{A}_i \mathcal{A}_j \mathcal{A}_k]$ above starting from the second Chern form $\mathrm{Tr}[\mathcal{F} \wedge \mathcal{F}]$; the relationship between the two is that

$$dK = \mathrm{Tr}[\mathcal{F} \wedge \mathcal{F}], \tag{2.36}$$

just as $\mathcal{A}$ is related to the first Chern form: $d(\mathrm{Tr}\mathcal{A}) = \mathrm{Tr}\mathcal{F}$. These relationships hold locally (this is known as Poincare's lemma, that given a closed form, it is *locally* an exact form) but not globally, unless the first or second Chern form generates the trivial cohomology class. For example, we saw that the existence of a nonzero first Chern number on the sphere prevented us from finding globally defined wavefunctions that would give an $\mathcal{A}$ with $d\mathcal{A} = \mathcal{F}$. We are assuming in even writing the Chern-Simons formula for θ that the ordinary Chern numbers are zero, so that an $\mathcal{A}$ can be defined in the 3D Brillouin zone. We would run into trouble if we assumed that an $\mathcal{A}$ could be defined in the 4D Brillouin zone if the *first or second* Chern number were nonzero. Note that the electromagnetic action above is just the second Chern form of the (Abelian) electromagnetic field.

The second Chern form is closed and hence generates an element of the de Rham cohomology we studied earlier. There are higher Chern forms as well: the key is that symmetric polynomials can be used to construct closed forms, by the antisymmetry properties of the exterior derivative. In physics, we typically keep the manifold fixed (in our Brillouin zone examples, it is usually a torus T^n), and are interested in classifying different fiber bundles on the manifold. In mathematical language, we want to use a properly normalized cohomology form to compute a homotopy invariant (i.e., with respect to changing the connection, not the manifold). This is exactly what we did with the Chern number in the IQHE, which was argued to compute certain integer-valued homotopy π_2 invariants of nondegenerate Hermitian matrices.

More precisely, we saw that the $U(1)$ gauge-dependence of polarization was connected to the homotopy group $\pi_1(U(1)) = \mathbb{Z}$, but that this is connected also to the existence of integer-valued Chern numbers, which we explained in terms of π_2. (These statements are not as inconsistent as they might seem, because our calculation of π_2 came down to π_1 of the diagonal unitary group.) We can understand the second Chern and Chern-Simons form similarly, using the homotopy invariants π_3 (gauge transformation in $d = 3$) and π_4 (quantized state in $d = 4$). The Chern-Simons integral for θ

	Polarization	Magnetoelectric polarizability
$d_{\min}$	1	3
Observable	$\mathbf{P} = \partial\langle H\rangle/\partial E$	$M_{ij} = \partial\langle H\rangle/\partial E_i \partial B_j$ $= \delta_{ij}\theta e^2/(2\pi h)$
Quantum	$\Delta\mathbf{P} = e\mathbf{R}/\Omega$	$\Delta M = e^2/h$
Surface	$q = (\mathbf{P}_1 - \mathbf{P}_2)\cdot\hat{\mathbf{n}}$	$\sigma_{xy} = (M_1 - M_2)$
EM coupling	$\mathbf{P}\cdot\mathbf{E}$	$M\mathbf{E}\cdot\mathbf{B}$
CS form	$\mathcal{A}_i$	$\epsilon_{ijk}(\mathcal{A}_i\mathcal{F}_{jk} + i\mathcal{A}_i\mathcal{A}_j\mathcal{A}_k/3)$
Chern form	$\epsilon_{ij}\partial_i\mathcal{A}_j$	$\epsilon_{ijkl}\mathcal{F}_{ij}\mathcal{F}_{kl}$

Table 2.1 Comparison of Berry-phase theories of polarization and magnetoelectric polarizability.

given above, in the non-Abelian case, has a $2\pi n$ ambiguity under gauge transformations, and this ambiguity counts the integer-valued homotopy invariant

$$\pi_3(SU(N)) = \mathbb{Z}, \quad N \geq 2. \tag{2.37}$$

In other words, there are "large" (non-null-homotopic) gauge transformations. Note that the Abelian Chern-Simons integral is completely gauge-invariant, consistent with $\pi_3(U(1)) = 0$.

The quantized state in $d = 4$ was originally discussed in the context of time-reversal-symmetric systems. The set $\mathcal{Q}$ has one integer-valued π_4 invariant for each band pair, with a zero sum rule. These invariants survive even once T is broken, but realizing the nonzero value requires that two bands touch somewhere in the four-dimensional Brillouin zone. In this sense, the "four-dimensional quantum Hall effect" is a property of how pairs of bands interact with each other, rather than of individual bands. Even if this 4D QHE is not directly measurable, it is mathematically connected to the 3D magnetoelectric polarizability in the same way as 1D polarization and the 2D IQHE are connected.

The above Chern-Simons formula for θ works, in general, only for a noninteracting electron system. This is not true for the first Chern formula for the IQHE, or the polarization formula, so what is different here? The key is to remember that the 3D Chern formula behaves very differently in the Abelian and non-Abelian cases; for example, in the Abelian case, θ is no longer periodic as the integral is fully gauge-invariant. Taking the ground state many-body wavefunction and inserting it into the Chern-Simons formula is not guaranteed to give the same result as using the multiple one-particle wavefunctions.

However, we can give a many-body understanding of θ that clarifies the geometric reason for its periodicity even in a many-particle system. Consider evaluating dP/dB by applying the 3D polarization formula

$$P_i = e\int_{BZ}\frac{d^3k}{(2\pi)^3}\mathrm{Tr}\,\mathcal{A}_i\,. \tag{2.38}$$

to a rectangular-prism unit cell. The minimum magnetic field normal to one of the faces that can be applied to the cell without destroying the periodicity is one flux quantum per unit cell, or a field strength $h/(e\Omega)$, where Ω is the area of that face. The ambiguity of polarization (2.38) in this direction is one charge per transverse unit cell area, i.e., e/Ω. Then the ambiguity in dP/dB is

$$\Delta\frac{P_x}{B_x} = \frac{e/\Omega}{h/(e\Omega)} = \frac{e^2}{h} = 2\pi\frac{e^2}{2\pi h}. \tag{2.39}$$

So the periodicity of 2π in θ is really a consequence of the geometry of polarization, and is independent of the single-electron assumption that leads to the microscopic Chern-Simons formula.

2.0.9 Anomalous Hall effect and Karplus-Luttinger anomalous velocity

Our previous examples of Berry phases in solids have concentrated on insulators, but one of the most direct probes of the Berry phase of Bloch electrons is found in metals that break time-reversal symmetry. The breaking of T allows a nonzero transverse conductivity σ_{xy} to exist along with the metallic diagonal conductivity σ_{xx}. This "anomalous Hall effect" (AHE) can originate from several different microscopic processes. The most interesting from a geometric point of view is the intrinsic AHE that results from Berry phases of a time-reversal-breaking band structure when the Fermi level is in the middle of a band. We will not attempt to discuss this interesting physics here but refer the reader to a recent comprehensive review [**?**] and note that there are an increasing number of other examples of Berry-phase effects in metals.

3
Introduction to topological order

These lectures are our first exposure to strongly interacting topological phases, defined as those that cannot be understood in terms of free particles. In contrast, the integer quantum Hall effect and topological insulators can be understood in terms of free particles, although these phases are stable in the sense that they survive over a finite region of interaction strength until a phase transition occurs. Our main tool will be quantum field theory, which is a powerful language to describe the long-wavelength physics of interacting systems. After giving some microscopic motivation from the fractional quantum Hall effect (FQHE), we give a first example of field theory applied to spin chains as an example of how an analysis of topological terms in a simple field theory led to a clear experimental prediction (the "Haldane gap") regarding antiferromagnetic integer-spin Heisenberg chains.

We then return to the quantum Hall effect and develop Abelian Chern-Simons theory, an example of a truly topological field theory. Although it is written in terms of one or more $U(1)$ gauge fields, similar to ordinary electromagnetism, its behavior is strikingly different than the conventional field theories with which the reader may already be familiar. In lieu of a microscopic derivation, which has been carried out but is somewhat tedious, we show that it unifies properties such as ground state degeneracy, braiding statistics, and edge excitations. We will follow increasingly standard parlance and use the term "topological order" specifically for phases of matter described by a non-trivial topological field theory, hence having ground state degeneracy, fractional statistics, etc. Thus the integer quantum Hall effect, which is certainly a topological phase of matter and well described by the Abelian Chern-Simons theory given below with $k = 1$, does not have topological order in the sense introduced by Wen.

3.1 FQHE background

We give quickly some standard background on the fractional quantum Hall effect in order to motivate the Chern-Simons field theory introduced below. The goal of that field theory is to give a compact universal description of the key features of the topological order in quantum Hall states, similar in spirit to the Ginzburg-Landau field theory of symmetry-breaking phases. Most of this material is standard and can be found in quantum Hall edited volumes and textbooks (Prange and Girvin; Das Sarma and Pinczuk; Jain).

Our discussion centers on the Laughlin wavefunction for two-dimensional electrons ($z_j = x_j + iy_j$ describes the jth electron, $j = 1, \ldots, N$)

$$\Psi_m = \left(\prod_{i<j} (z_i - z_j)^m \right) e^{-\sum_i |z_i|^2/4\ell^2}. \tag{3.1}$$

The magnetic length is $\ell = \sqrt{\hbar c/eB}$ and the wavefunction is not normalized. This wavefunction clearly can be expanded over the single-electron lowest Landau level wavefunctions in the rotational gauge,

$$\psi_m = z^m e^{-|z|^2/4\ell^2}. \tag{3.2}$$

where $m = 0, 1, \ldots$ labels angular momentum. For $m = 1$ the Laughlin state is just a Slater determinant for the filled lowest Landau level, but for higher m it is believed not to be a sum of any finite number of Slater determinants in the $N \to \infty$ limit.

This wavefunction can be justified using the pseudopotential approach introduced by Haldane: it is the maximum-density zero-energy state of a repulsive interaction that vanishes for relative angular momentum greater than or equal to m. We checked that its density is $\nu = 1/m$ by looking at the degree of the polynomial factor, which is directly related to $\langle r^2 \rangle$, and argued that it contains "quasihole" excitations of charge $-q/m$, where q is the charge of the electrons. The wavefunction for a quasihole at z_0 is

$$\Psi_{\text{quasihole}} = \left(\prod_i (z_i - z_0) \right) \Psi_m. \tag{3.3}$$

The fractional charge can be understood by noting that m copies of the extra factor here would lead to the wavefunction with an electron at z_0, but without treating z_0 as an electron coordinate; in other words, a wavefunction with a "hole" added at z_0. It has edge states that at first glance are loosely similar to those in the filled Landau level.

3.2 Topological terms in field theories: the Haldane gap and WZW models

As a warm-up for fully topological field theories, we give an example of how topological terms can have profound consequences in "ordinary" field theories (i.e., theories without gauge fields). By a topological term we mean loosely one whose value in any specific configuration (a path in the path integral) is a topological invariant, so that the set of all paths can be divided into topological sectors by the value of the topological term. A famous example of this phenomenon found by Haldane led to the first understanding of the gapped spin-one Heisenberg antiferromagnet in one spatial dimension, which has recently been interpreted [**?**, **?**] as a symmetry-protected trivial (SPT) phase of interacting particles. We will focus on topological terms that appear in nonlinear σ-models, which despite their unwieldy name are a very basic type of field theory for systems in or near an ordered phase breaking a continuous symmetry.

We first present Haldane's example (following closely the treatment of Auerbach [**?**]), and then discuss a different kind of topological term that appears in Wess-Zumino-Witten models, again in one spatial dimension; details of the latter are provided in

an appendix. For higher dimensions, only a few general comments are provided at the end, but we return to the subject when we consider nonlinear σ-models for disordered systems in section **??**. The nonlinear σ-model (NLSM) is an example of an effective theory, a simplified description of the low-energy degrees of freedom of a complicated system. Ginzburg-Landau theory is another such effective theory, and one use of the NLSM is as a further simpification of Ginzburg-Landau theory where we have thrown away the "hard" or "massive" fluctuations of the magnitude of the order parameter, keeping only the "soft" or "massless" fluctuations within the order parameter manifold.

For definiteness, consider a d-dimensional XY model, which would be described in Ginzburg-Landau theory by a 2-component real or 1-component complex order parameter. The mean-field physics in the ordered phase as a function of the order parameter is illustrated in Fig. **??**: the order parameter manifold of symmetry-related ground states is a circle, and we can expect that fluctuations along this circle are "soft" in the sense of requiring little energy (as this is a flat direction of the energy) while those perpendicular to the circle are more costly. This order parameter manifold is the same as that considered in the discussion of topological defects in Section **??**, where defects were classified using maps from surfaces enclosing the defect in real space to the order parameter manifold. At low temperature we might expect that a reasonable description of the system is therefore obtained just from fluctuations of the order parameter's direction, leading to a functional integral for the coarse-grained classical partition function:

$$Z_{\mathrm{NLSM}} = \int \mathcal{D}\theta(x) e^{-\beta c \int \frac{(\nabla\theta)^2}{2} d^d x}. \tag{3.4}$$

Here c is a coupling constant with units of energy if $d = 2$; one could estimate c simply from the coupling strength in a lattice XY model. The NLSM it is called nonlinear because the circle is defined by a hard constraint on the $\hat{\mathbf{n}}$ field, which in more complicated target manifolds such as the sphere leads to interaction (i.e., nonlinear) terms in the fields obtained in a perturbative expansion; it is called a sigma model because of its first appearance in particle physics.

For a quantum-mechanical model at zero temperature, we might expect on general grounds that imaginary time will become an extra dimension in any Euclidean path-integral representation of the partition function, in the same way as the Dirac-Feynman path integral for quantum mechanics involves integration of the Lagrangian over time (a 0+1-dimensional theory). Now we will obtain a NLSM for a quantum-mechanical problem in one spatial dimension. Heuristically, we might expect an NLSM to be a reasonable description for a quantum model that is "close to" having symmetry-breaking order.

Our approach is to derive a connection between the low-energy, long-wavelength degrees of freedom of the spin path integral of the Heisenberg antiferromagnet. This process is known as Haldane's mapping in the context of spin systems: we will use it to show that there is a topological term present for half-integer spin but not for integer spin, which is believed to explain the different behavior seen numerically and experimentally in these two cases.

First we look for a more general way of writing the Berry phase term for a spin that results from setting up a coherent-state path integral for spin. In order to make a path integral, we should set up an integral over "classical" trajectories; what is the classical trajectory of a spin? One answer is to use the overcomplete basis of coherent states for the spin-S Hilbert space (see the book of Auerbach), which are labeled by a unit vector $\hat{\Omega}$. As S increases, the spin wavefunction becomes more and more concentrated around $\hat{\Omega}$.

$$\omega[\hat{\Omega}] = -\int_0^\beta d\tau \dot{\phi} \cos\theta. \tag{3.5}$$

For a closed path on the sphere, we showed in Chapter II that this corresponds to the signed spherical area enclosed by the path. An overall ambiguity of $\pm 4\pi$ in this area does not affect the physics, since the area ω appears in the path integral action with a coefficient $-iS$. For a many-spin system, the full action was

$$S[\hat{\Omega}] = -iS \sum_i \omega[\hat{\Omega}_i] + \int_0^\beta d\tau \, \frac{S^2 J}{2} \sum_{ij} \hat{\Omega}_i \cdot \hat{\Omega}_j. \tag{3.6}$$

For now we return to a single spin to set up an improved way of writing the Berry phase term.

Let the vector potential $\mathbf{A}(\hat{\Omega})$ be assumed to have the following property: its line integral over a closed orbit on the sphere should give the area enclosed by the orbit,

$$\omega = \int_0^\beta d\tau \, \mathbf{A}(\hat{\Omega}) \dot{\hat{\Omega}}. \tag{3.7}$$

Then Stokes's theorem fixes curl $\mathbf{A}$ to be the magnetic field of a magnetic monopole (a vector with uniform outward component):

$$\nabla \times \mathbf{A} \cdot \hat{\Omega} = \epsilon^{\alpha\beta\gamma} \frac{\partial A_\beta}{\partial \hat{\Omega}_\alpha} \hat{\Omega}^\gamma = 1. \tag{3.8}$$

Two explicit examples to check that this can be done are

$$\mathbf{A}^a = -\frac{\cos\theta}{\sin\theta} \hat{\phi}, \quad \mathbf{A}^b = \frac{1-\cos\theta}{\sin\theta} \hat{\phi}. \tag{3.9}$$

Clearly $\mathbf{A}^a$ has singularities at the north and south poles, while $\mathbf{A}^b$ has a singularity only at the south pole. (Parenthetical note: actually $\mathbf{A}^b$ is a good representation of the field of a Dirac monopole: a singular flux ("Dirac string") enters through the south pole, and then goes out uniformly over the rest of the sphere. A small circle around the south pole contains flux 4π, which contributes $4\pi S$ to the action, but recall that this winds up giving zero physical contribution to the path integral.)

Now we can use this representation to write concisely the variation of the Berry phase term under a small variation in the path from imaginary time 0 to imaginary time t. Suppose that we want to calculate

$$\delta\omega[\hat{\Omega}] = \int_0^t dt' \, \delta(\mathbf{A} \cdot \dot{\hat{\Omega}}) = \int_0^t dt' \, (\frac{\partial A^\alpha}{\partial \hat{\Omega}^\beta} \delta\hat{\Omega}^\beta \dot{\hat{\Omega}}^\alpha + A^\alpha \frac{d}{dt} \delta\hat{\Omega}^\alpha) \tag{3.10}$$

under a small variation of the path $\delta\hat{\Omega}$ that is assumed to keep the endpoints fixed. Now subtract $\frac{\partial A^\alpha}{\partial \hat{\Omega}^\beta}\dot{\hat{\Omega}}^\beta \delta\hat{\Omega}^\alpha$ from the first term and add it to the second, to get

$$\begin{aligned}\delta\omega[\hat{\Omega}] &= \int_0^t dt' \, \frac{\partial A^\alpha}{\partial \hat{\Omega}^\beta}\epsilon^{\alpha\beta\gamma}(\dot{\hat{\Omega}}\times\delta\hat{\Omega})_\gamma + \int_0^t dt' \, (A^\alpha \frac{d}{dt}\delta\hat{\Omega}^\alpha + \frac{\partial A^\alpha}{\partial \hat{\Omega}^\beta}\dot{\hat{\Omega}}^\beta \delta\hat{\Omega}^\alpha) \\ &= \int_0^t dt' \, \hat{\Omega}\cdot(\dot{\hat{\Omega}}\times\delta\hat{\Omega}) + \int_0^t dt' \, \frac{d}{dt'}(\mathbf{A}\cdot\delta\hat{\Omega}) = \int_0^t dt' \, \hat{\Omega}\cdot(\dot{\hat{\Omega}}\times\delta\hat{\Omega}). \quad (3.11)\end{aligned}$$

Here we used the condition (3.8) and also, in rewriting the first term, the fact that the quantity in parentheses $(\dot{\hat{\Omega}}\times\delta\hat{\Omega}) \parallel \hat{\Omega}$ because of the constant length of the vector $\hat{\Omega}$.

Now, after this prelude, we are ready to rewrite the full path integral for the many-spin Heisenberg model. The first step is to write the spin $\hat{\Omega}_i$ in terms of two continuous fields of spacetime $\hat{\mathbf{n}}$ and $\mathbf{L}$:

$$\hat{\Omega}_i = \eta_i \hat{\mathbf{n}}(\mathbf{x}_i)\sqrt{1-\left|\frac{\mathbf{L}(\mathbf{x}_i)}{S}\right|^2} + \frac{\mathbf{L}(\mathbf{x}_i)}{S}. \quad (3.12)$$

Here η_i alternates between sublattices, $\hat{\mathbf{n}}(\mathbf{x})$ is a unit vector field, sometimes referred to as the Néel field, and $\mathbf{L}$ is constrained to be perpendicular to $\hat{\mathbf{n}}$. Hence a constant value of $\hat{\mathbf{n}}$ corresponds to a classical Néel state. It seems like we have greatly increased the degrees of freedom by this rewriting; what we do now is restrict the allowed Fourier components of the new fields (i.e., the Brillouin zone) to small momenta in such a way that the total number of degrees of freedom is unchanged (cf. Auerbach for details). The spirit of this approximation is that we are interested in long-length-scale physics so details on the scale of the lattice spacing are unimportant. It turns out that we assume slow variations in $\hat{\mathbf{n}}$ but only that $|\mathbf{L}| \ll S$, i.e., that $\mathbf{L}$ is small but not necessarily slowly varying. We now expand the path integral in powers of $|\mathbf{L}|/S$.

A pair of spins gives a contribution

$$\hat{\Omega}_i\cdot\hat{\Omega}_j \approx \eta_i\eta_j - \frac{1}{2}\eta_i\eta_j(\hat{\mathbf{n}}_i-\hat{\mathbf{n}}_j)^2 + \frac{1}{S^2}\left[\mathbf{L}_i\mathbf{L}_j - \frac{1}{2}\eta_i\eta_j(\mathbf{L}_i^2+\mathbf{L}_j^2)\right] + \frac{1}{S}(\eta_j\mathbf{L}_i\hat{\mathbf{n}}_j + \eta_i\mathbf{L}_j\hat{\mathbf{n}}_i) + \ldots \quad (3.13)$$

Here the neglected terms are of order $|\mathbf{L}|^2(\hat{\mathbf{n}}_i-\hat{\mathbf{n}}_j)$ or smaller. In the first term, use a Taylor expansion to convert differences of the Néel field into derivatives and keep only the leading contribution. You can show (or it is in Auerbach) that the cross terms (those with both $\mathbf{L}$ and $\hat{\mathbf{n}}$) vanish by the symmetry of the Heisenberg Hamiltonian. The term with two $\mathbf{L}$ factors we rewrite below in Fourier space, where it is much simpler and where we will be able to "integrate it out".

What we are left with, after going from the lattice to integrals using

$$\sum_i F_i \to a^{-d}\int d^dx \, \sum_i \delta(\mathbf{x}-\mathbf{x}_i)F(x), \quad (3.14)$$

is the continuum representation

$$H = E_0 + \frac{1}{2}\int d^dx \left[\rho_s \sum_l |\partial_l\hat{\mathbf{n}}|^2 + \int d^dx' \, (\mathbf{L}_x\chi^{-1}_{xx'}\mathbf{L}_{x'})\right]. \quad (3.15)$$

Here E_0 is the classical energy

$$E_0 = \frac{S^2}{2} \sum_{ij} J_{ij} \eta_i \eta_j. \tag{3.16}$$

In the first term, the spin stiffness is

$$\rho_s = -\frac{S^2}{2dNa^d} \sum_{ij} J_{ij} \eta_i \eta_j |\mathbf{x}_i - \mathbf{x}_j|^2. \tag{3.17}$$

The second or "canting" term in Fourier space is simply

$$\int \frac{d^d q}{(2\pi)^d} \frac{\mathbf{L_q} \mathbf{L_{-q}}}{J(\mathbf{q}) - J(\pi, \pi, \ldots)}, \quad J(\mathbf{q}) = \sum_j e^{i\mathbf{q}\cdot(\mathbf{x}_i - \mathbf{x}_j)} J_{ij}. \tag{3.18}$$

Now we just need to rewrite the geometric phase

$$-iS \sum_i \omega_i = -iS \int_0^\beta d\tau \sum_i \mathbf{A}(\hat{\Omega}_i) \dot{\hat{\Omega}}_i. \tag{3.19}$$

Assume that the vector potential is chosen so that $\mathbf{A}(\hat{\Omega}) = \mathbf{A}(-\hat{\Omega})$, as works for one of the examples above. Now expanding in terms of the new fields,

$$\begin{aligned} -iS \sum_i \omega_i &= -iS \sum_i \eta_i \omega[\hat{\mathbf{n}}_i + \eta_i \mathbf{L}_i / S] \\ &= -iS \sum_i \left[\eta_i \omega[\hat{\mathbf{n}}_i + \frac{\delta \omega}{\delta \hat{\mathbf{n}}_i} \cdot (\mathbf{L}_i / S) \right] \\ &= -i\Upsilon - i \int_0^\beta d\tau \sum_i (\hat{\mathbf{n}}_i \times \partial_\tau \hat{\mathbf{n}}_i \cdot \mathbf{L}_i). \end{aligned} \tag{3.20}$$

In the last line we used our earlier formula for the variation of ω, and defined

$$\Upsilon = S \sum_i \eta_i \omega[\hat{\mathbf{n}}(\mathbf{x}_i)], \tag{3.21}$$

switching to the spatial continuum limit.

Now our goal is going to be to combine the classical and geometric terms in order to obtain a simple long-wavelength action. The key step is to note that the second term in (3.20) couples one power of L to a combination of n fields. So integrating out the L degrees of freedom (a Gaussian integral) will give rise to the following: considering only the terms involving $\mathbf{L}$ and doing the integral in Fourier space, we get (ignoring an unimportant overall constant)

$$Z_L \propto \int \mathcal{D}\hat{\mathbf{n}} \exp\left[-\frac{1}{2} \int d\tau \frac{d^d q}{(2\pi)^d} (J(\mathbf{q}) - J(\vec{\pi}))(\hat{\mathbf{n}} \times \partial_\tau \hat{\mathbf{n}})_{\mathbf{q}} \cdot (\hat{\mathbf{n}}_\times \partial_\tau \hat{\mathbf{n}})_{-\mathbf{q}} \right]. \tag{3.22}$$

We can simplify this much further: for long wavelengths we approximate $\chi(\mathbf{q}) \approx \chi(0)$, and use

$$|(\hat{\mathbf{n}} \times \partial_\tau \hat{\mathbf{n}}|^2 = |\partial_\tau \hat{\mathbf{n}}|^2 \tag{3.23}$$

from the constant length of $\hat{\mathbf{n}}$. to get just (the real-space constant $\chi_0 = a^{-d}\chi(0) = a^{-d}(J(\vec{0}) - J(\vec{\pi}))$)

$$Z_L = \int \mathcal{D}\hat{\mathbf{n}} \exp\left(-\frac{1}{2}\int_0^\beta d\tau \int d^d x\, \chi_0 |\partial_\tau \hat{\mathbf{n}}|^2\right). \tag{3.24}$$

So, putting it all together, we have

$$Z \propto \int \mathcal{D}\hat{\mathbf{n}}\, e^{i\Upsilon} \exp\left[-\frac{1}{2}\int_0^\beta d\tau \int d^d x \left(\chi_0 |\partial_\tau \hat{\mathbf{n}}|^2 + \rho_s |\partial_{x^\alpha} \hat{\mathbf{n}}|^2\right)\right]. \tag{3.25}$$

This now looks much more symmetric between space and time; if desired, one can just rescale time to make the theory look like it lives in an isotropic $d+1$-dimensional space. This gives

$$Z \propto \int \mathcal{D}\hat{\mathbf{n}}\, e^{i\Upsilon} \exp(-\int d^{d+1}x \mathcal{L}_{NLSM}), \quad \mathcal{L}_{NLSM} = \sum_{\alpha=1}^{d+1} \frac{\partial_{x^\alpha}\hat{\mathbf{n}} \cdot \partial_{x^\alpha}\hat{\mathbf{n}}}{2}. \tag{3.26}$$

This NLSM is the simplest field theory of maps from the space $\mathcal{R}^{d+1}$ to the unit sphere. We still need to say a bit about the topological term Υ (the capital Greek letter upsilon): in one spatial dimension this term fundamentally modifies the physics, for reasons we shall see. We expand it for slowly varying $\hat{\mathbf{n}}(x)$: recall that Υ is defined to include factors η_i, so

$$\Upsilon^{d=1} = -S\sum_i (\omega[\hat{\mathbf{n}}(x_{2i})] - \omega[\hat{\mathbf{n}}(x_{2i-1})]) = \frac{S}{2}\int \frac{dx}{a}\frac{\delta\omega}{\delta\hat{\mathbf{n}}} \cdot \partial_x \hat{\mathbf{n}} a = 2\pi S \Theta[\hat{\mathbf{n}}(x,\tau)]. \tag{3.27}$$

Here Θ comes from using our previous variation form for the variation $d\omega$:

$$\Theta = \frac{1}{4\pi}\int d\tau \int dx\, (\hat{\mathbf{n}} \times \partial_\tau \hat{\mathbf{n}} \cdot \partial_x \hat{\mathbf{n}}). \tag{3.28}$$

This form is known as the Pontryagin index, which is a topological invariant like a winding number. It is an integer and is constant under smooth deformations of $\hat{\mathbf{n}}$. Essentially it measures the number of times the map from $(-L/2, L/2)\times(0,\beta)$ "wraps" the sphere S^2. You can easily construct examples with $\Theta = 0$ (the constant map) and $\Theta = 1$ (spherical projection). If you want a sense for why it is a topological invariant (which is not that hard to show), imagine that someone gives you a sphere wrapped with paper. The paper can't be "contracted to a point" without tearing, unlike a loop drawn on the sphere. So maps $S^1 \to S^2$ are all contractible, while maps $S^2 \to S^2$ are classified by the Pontryagin index.

The important thing to note is that the coefficient in front of this integer is only $2\pi S$, so that there will be a difference between integer and half-integer spins. For integer spin the topological term doesn't do anything, while for half-integer spins,

there is interference between terms with odd or even values of the Pontryagin index. So Haldane's mapping explains (with a few approximations along the way!) the profound difference between integer and half-integer spins in one dimension, later confirmed experimentally and numerically. It is actually easier just to solve the spin-half chain using the Bethe ansatz than to explicitly solve its continuum theory with Berry phases, although a proof has been given that the latter is indeed gapless. The Lieb-Schultz-Mattis theorem discussed by Chalker elsewhere in this volume provides a general reason why the half-integer-spin case is gapless. Experimental results by neutron scattering (cf. Nagler et al.) confirm the existence of a gap and also the existence of spin-half edge states at the end of chains, which are discussed elsewhere in this volume by Regnault.

So we have seen how the unusual geometry of spin space, in the path-integral representation, gives rise to a profound difference between integer-spin and half-integer spin chains. Further reading and references for such topics can be found in the book of Auerbach. We can connect the above result to the exact solution by Affleck, Kennedy, Lieb, and Tasaki (AKLT) of a spin-1 chain with additional biquadratic interactions:

$$H = J \sum_i \left[S_i \cdot S_j + \alpha (S_i \cdot S_j)^2 \right] \tag{3.29}$$

with $J > 0$ and $\alpha = 1/3$. This value of α is special in that the two terms on each bond act as projectors onto the total spin-0 and spin-1 subspaces of the two spins, with equal weight. The full phase diagram of the bilinear-biquadratic phase diagram from numerical density-matrix renormalization group studies is shown in Fig. 3.1. We note that the Haldane problem of the purely bilinear chain is in the same gapped phase as the AKLT solution, but that there are other phases as well, and there are also parameter values for which the system is gapless. As the last part of our discussion of topological terms for now, we explain the meaning of the labels $SU(3)_1$ and $SU(2)_2$ on the critical points in Fig. 3.1, which combine a Lie group with a subscripted integer known as the level. These points are examples of field theories with both conformal invariance and Lie group symmetry known as Wess-Zumino-Witten (WZW) models.

The NLSM for the XY model in equation (3.30) can be written in a different way if we think about the order parameter manifold (the circle) as the Lie group $U(1)$. Writing $g = e^{i\theta}$, we note that $\partial_i \theta = g^{-1} \partial_i g$, so

$$Z_{\mathrm{NLSM}} = \int \mathcal{D}\theta(x) e^{-\beta c \int d^2x \, \sum_i (g^{-1}\partial_i g)^2/2)}. \tag{3.30}$$

In taking the trace here, we are looking ahead to a generalization. There is not a Lie group structure on the sphere, but we might be tempted to generalize to other Lie groups, for example by taking $g \in U(N)$ or $SU(N)$. Then $g^{-1}\partial_i g$ is an element of the Lie algebra, which has an inner product known as the Killing form; for $SU(N)$, $\mathcal{K}(X, Y) = 2N\mathrm{Tr}(XY)$. Generalizing the kinetic term that is the only term in the action above to the Lie algebra is straightforward.

However, it turns out that the low-energy physics of this generalization with just the resulting term is quite different than the $U(1)$ case. As for the NLSM into the sphere, the fact that the manifolds of unitary groups are curved once we go beyond the circle, leading to interactions that result in a mass gap. If we want instead to obtain

a gapless model with Lie group symmetry, we must add an additional topological term first written down by Wess and Zumino. This term is quite unusual in that it requires extending the manifold on which the theory lives into an extra dimension. Assume $N > 1$ in what follows, and pick $g \in SU(N)$ for definiteness. Let us compactify the two-dimensional space into S^2 as for the Haldane chain above. Given a configuration of the Lie-group field g on the surface of a sphere, we can always find a way to smoothly deform that configuration to the constant configuration since $\pi_2(SU(N))$ is trivial.

We will keep writing the generalized model in Euclidean space although their primary relevance is to quantum models in one spatial dimension. The action of the Wess-Zumino-Witten model in the usual notation is then (see Appendix for a physics motivation)

$$S = -\frac{k}{8\pi}\int_{S^2} d^2x\, \mathcal{K}(g^{-1}\partial^\mu g, g^{-1}\partial_\mu g) - \frac{k}{24\pi}\int_{B^3} d^3y\, \epsilon^{ijk}\mathcal{K}(g^{-1}\partial_i g, [g^{-1}\partial_j g, g^{-1}\partial_k g]). \tag{3.31}$$

The meaning of upper and lower indices in the first term is that the metric of spacetime appears. In the second term, in contrast, the ϵ term appears instead of the metric, a sign that the term is topological in the sense of being metric-independent. In the second term, we have chosen a continuation of the field g into the interior B^3 of the sphere S^2. While as mentioned above those continuations certainly exist, we should check to make sure that the physics is independent of precisely which continuation we chose.

Fig. 3.1 Phase diagram of the bilinear-biquadratic spin-1 chain from Ref. [**?**]. The Hamiltonian is $H = J\sum_i \left[\cos\theta S_i \cdot S_j + \sin\theta (S_i \cdot S_j)^2\right]$, with $J > 0$.

This independence is related to another topological fact about $SU(N)$. Consider two different continuations from S^2 into B^3. Actually, as a simpler example, consider two different continuations from S^1 into B^2. We could combine those into a field configuration on S^2, where one continuation gives the northern hemisphere and the other gives the southern hemisphere. In the same way, combining our two continuations from S^2 to B^3 gives a field configuration on S^3. Since $\pi_3(SU(N)) = \mathbb{Z}$, there are integer-valued classes of such configurations, and in fact the Wess-Zumino term is defined so as to compute this topological invariant Z: more precisely, the difference of the integral above for two different continuations into the bulk is k times $2\pi n$, where $n \in \mathbb{Z}$ measures the topological invariant of the map $S^3 \to S^3$ resulting from combining the two continuations as described above.

When we put this action into a quantum path integral, it therefore leads to a quantization of the level k to *integers*. $SU(2)_k$ with $k = 2$ can be viewed as a different representation of the same symmetry as the $SU(2)_1$ realized in the spin-half Heisenberg chain, in the same way as the spins on one site are in different representations of ordinary $SU(2)$. The full demonstration that the model is gapless is beyond our present scope, but at least we have a topological understanding of why the Wess-Zumino term

is a natural quantity to consider. One way to tell apart the gapless points associated with different levels or Lie groups is by computing the central charge c, a measure of how many degrees of freedom are gapless at the critical point, in units where one free boson gives $c = 1$. The WZW model for Lie group g at level k has central charge

$$c = \frac{k \dim SU(N)}{k+n} \tag{3.32}$$

where $\dim SU(N) = N(N-1)$. Hence $SU(3)_1$ has central charge 2, $SU(2)_2$ has $c = 3/2$, and $SU(2)_1$ has $c = 1$, consistent with its bosonized representation as a single boson.

3.3 Topologically ordered phases: the fractional quantum Hall effect

3.3.1 Chern-Simons theory I: flux attachment and statistics change

We will now start the process of developing a more abstract description of the fractional quantum Hall effect that will help us understand what type of order it has. For example, this will define precisely what it means to say that the physical state is adiabatically connected to the Laughlin wavefunction. Our main tool will be Chern-Simons theory; we briefly encounted the Chern-Simons term of the electromagnetic gauge potential when we discussed quantum Hall layers at the surface of the strong topological insulator, and we will come to that in a moment. However, a more fundamental use of the Chern-Simons theory is to describe the internal degrees of freedom of the quantum Hall liquid. In other words, we will have both an "internal" Chern-Simons theory describing the quantum Hall liquid and a Chern-Simons term induced in the electromagnetic action.

Since that sounds complicated, let's start by understanding why a Chern-Simons theory might be useful. To begin, we come up with a picture for the Laughlin state by noting that, since the filled lowest Landau level has one magnetic flux quantum per electron, the Laughlin state at $m = 3$ (i.e., $\nu = 1/3$) has three flux quanta per electron. To get a picture for how the Laughlin state is connected to the $\nu = 1$ state, we imagine attaching two of these flux quanta to each electron. The resulting "composite fermion" still has fermionic statistics, by the following counting. Interchanging two electrons gives a -1 factor. The Aharonov-Bohm factor from moving an electron all the way around a flux quantum is $+1$, but in this exchange process, each electron moves only half-way around the flux quanta attached to the other electron. So when one of these objects is exchanged with another, the wavefunction picks up three factors of -1 and the statistics is still fermionic.

These composite fermions now can form the integer quantum Hall state in the remaining field of one flux quantum per composite fermion, leading to a $\nu = 1/3$ incompressible state in terms of the original electrons. More generally, the phase picked up by a particle of charge q moving completely around a flux Φ is

$$e^{i\theta} = e^{iq\Phi/(\hbar c)}. \tag{3.33}$$

We will now see how the Chern-Simons term lets us carry out a "flux attachment" related to the above composite fermion idea: in fact, by attaching three flux quanta

rather than two to each electron, we would obtain bosons moving in zero applied field, and the Laughlin state can be viewed as a Bose-Einstein condensate of these "composite bosons" [**?**]. [1]

The Abelian Chern-Simons theory we will study is described by the Lagrangian density in 2+1 dimensional Minkowski spacetime

$$\mathcal{L} = 2\gamma\epsilon^{\mu\nu\lambda}a_\mu\partial_\nu a_\lambda + a_\mu j^\mu \tag{3.34}$$

where γ is a numerical constant that we will interpret later, a is the Chern-Simons gauge field, and j is a conserved current describing the particles of the theory. Under a gauge transformation $a_\mu \to a_\mu + \partial_\mu\chi$, the Chern-Simons term (the first one) transforms as

$$\epsilon^{\mu\nu\lambda}a_\mu\partial_\nu a_\lambda \to \epsilon^{\mu\nu\lambda}a_\mu\partial_\nu a_\lambda + \epsilon^{\mu\nu\lambda}\partial_\mu\chi\partial_\nu a_\lambda, \tag{3.35}$$

where the term with two derivatives of χ drops out by antisymmetry. The new term can be written as

$$\delta S = 2\gamma\int d^2x dt\epsilon^{\mu\nu\lambda}\partial_\mu(\chi\partial_\nu a_\lambda), \tag{3.36}$$

where again the term with two derivatives acting on a gives zero by antisymmetry. So, *if we can neglect the boundary*, the Abelian Chern-Simons term is gauge-invariant. (As we discussed previously in the discussion of magnetoelectric polarizability, the non-Abelian Chern-Simons term is not gauge-invariant, because "large" (non-null-homotopic) gauge transformations change the integral; this is related to the third homotopy group of $SU(N)$.) Later on we will actually consider a system with a boundary and see how the boundary term leads to physically important effects.

Consider the equation of motion from varying this action. We get

$$4\gamma\epsilon^{\mu\nu\lambda}\partial_n u a_\lambda = -j^\mu. \tag{3.37}$$

where the 4 rather than 2 appears because the Chern-Simons term has nonzero derivative with respect to both a and ∂a. For a particle sitting at rest, the spatial components of the current vanish, but there must be a flux: writing in components,

$$\int d^2x(\partial_1 a_2 - \partial_2 a_1) = -\frac{1}{4\gamma}\int d^2x\, j^0. \tag{3.38}$$

Hence a charged particle in the theory gains a flux of the a field (since the left term is just the integral of a magnetic field). If the charge is localized, then the flux is localized as well.

What good is this? Well, we know that when one charged particle with respect to the a field moves around another, it will now pick up an Aharonov-Bohm phase from the attached flux in addition to any statistics factor. The additional statistics factor is

$$\theta = \frac{1}{8\gamma}, \tag{3.39}$$

where the $1/2$ here results because the particles only move halfway around each other in an exchange. In other words, if we started with $\theta = 0$ bosonic particles but added

[1] One feature of the composite fermion picture that is preferable to the composite boson picture is that the former is naturally described as "topological order", while the latter would lead to a picture of the phase in terms of the symmetry-breaking order of a BEC.

a $\gamma = \frac{1}{8\pi}$ Chern-Simons term, we would obtain fermions, and vice versa. But so far nothing constrains γ, suggesting that in two dimensions, "braiding" statistics is not constrained to be bosonic or fermionic. Particles in two dimensions that are neither bosonic nor fermionic are known as "anyons".

Why is two spatial dimensions so special? It turns out that an argument about why generalized statistics are possible for point particles in two spatial dimensions but not higher dimensions was given long ago by Leinaas and Myrheim (1976). The key observation is that an exchange path that takes one particle around another and back to its original location is not smoothly contractible in 2D without having the particles pass through each other, while in higher dimensions, such a path is contractible. The consequence of this is that in two dimensions, phase factors are not just defined for permutations of the particles but rather for any "braiding". [2]

3.3.2 Chern-Simons theory II: integrating out gauge fields and coupling to electromagnetism

Aside from the composite fermion/composite boson pictures, why might the Chern-Simons theory with Lagrangian density given by (3.34) describe quantum Hall states? Without working through a detailed derivation starting from nonrelativistic quantum mechanics of many interacting electrons in a magnetic field (which is still not all that rigorous; for a discussion, see lecture notes of A. Zee in *Field Theory, Topology, and Condensed Matter Physics*, Springer), we can note the following. A conserved electromagnetic current in 2+1D can always be written as the curl of a gauge field:

$$J^\mu = \frac{1}{2\pi}\epsilon^{\mu\nu\lambda}\partial_\nu a_\lambda. \tag{3.40}$$

(Note that this electromagnetic current might in general be distinct from the particle current above.) Here a is automatically a gauge field since the $U(1)$ gauge transformation does not modify the current. Gauge invariance forbids the mass term $a^\mu a_\mu$, so the lowest-dimension possible term is the Chern-Simons term, which we write for future use with a different normalization than above:

$$\mathcal{L}_{CS} = \frac{k}{4\pi}\epsilon^{\mu\nu\lambda}a_\mu\partial_\nu a_\lambda. \tag{3.41}$$

The point of the new normalization $k = 8\pi\gamma$ compared to (3.34) is that the boson-fermion statistics transformation above now corresponds just to $k = 1$. We will argue later that k should be an integer for the electron to appear somewhere in the spectrum of excitations of the theory.

Does this term need to appear? No, for example, in a system that has P or T symmetry, it cannot appear. However, if it does appear, then since there is only one spatial derivative, it dominates the Maxwell term at large distances. Effectively we define the quantum Hall phase as one in which $\mathcal{L}_{CS}$ appears in the low-energy Lagrangian; for example, this is true in both the Laughlin state and the physical state with Coulomb

[2] Even non-Abelian statistics are possible if there are multiple ground states: the phase factor associated with a particular braid is then a matrix acting on the set of ground states, and two such matrices need not commute.

interactions, even though the overlap between those two ground-state wavefunctions is presumably zero in the thermodynamic limit.

What if we added the $a_\mu J^\mu$ coupling and integrated out the gauge field? Well, the main reason not to do that is that we obtain a nonlocal current-current coupling. Since the original action is quadratic in the fields, this integration is not too difficult, but an alternate, equivalent way to do it is to solve for a in terms of J. Given a general Lagrangian

$$\mathcal{L} = \phi \mathcal{Q} \phi + \phi J, \tag{3.42}$$

where $\mathcal{Q}$ denotes some operator, we have the formal equation of motion from varying ϕ

$$2\mathcal{Q}\phi = -J \tag{3.43}$$

which is solved by

$$\phi = \frac{-1}{2\mathcal{Q}} J. \tag{3.44}$$

Then substituting this into the Lagrangian (and ignoring some subtleties about ordering of operators), we obtain

$$\mathcal{L} = \frac{1}{4} J \frac{1}{\mathcal{Q}} J - J \frac{1}{2\mathcal{Q}} J = -J \frac{1}{4\mathcal{Q}} J. \tag{3.45}$$

So for the Chern-Simons term we need to define the inverse of the operator $\epsilon^{\mu\nu\lambda}\partial_\nu$ that appears between the a fields. This is a bit subtle because there is a zero mode of the original operator, related to gauge-invariance: for any smooth function g, $\epsilon^{\mu\nu\lambda}\partial_\nu(\partial_\lambda g) = 0$. To define the inverse, we fix the Lorentz gauge $\partial_\mu a_\mu = 0$. In this gauge, we look for an inverse using

$$(\epsilon^{\mu\nu\lambda}\partial_\nu)(\epsilon^{\lambda\alpha\beta}\partial_\alpha a_\beta) = \epsilon^{\mu\nu\lambda}\epsilon^{\lambda\alpha\beta}(\partial_\nu \partial_\alpha a_\beta). \tag{3.46}$$

We can combine the ϵ tensors by noting that $\epsilon^{\mu\nu\lambda} = \epsilon^{\lambda\mu\nu}$, so there are two types of nonzero terms in the above: either $\mu = \alpha$ and $\nu = \beta$ or vice versa, with a minus sign in the second case. From the first type of term, we obtain $\partial_\alpha(\partial_\beta a_\beta)$ which is zero by our gauge choice. From the second type, we obtain

$$-\partial_\nu^2 a_\mu. \tag{3.47}$$

So the inverse of the operator appearing in the Chern-Simons term in this gauge is $-\epsilon^{\mu\nu\lambda}\partial_\nu/\partial^2$, and the Lagrangian (3.34) with the gauge field integrated out is just

$$\mathcal{L} = \frac{1}{8\gamma} j_\mu \left(\frac{\epsilon^{\mu\nu\lambda}\partial_\nu}{\partial^2} \right) j_\lambda. \tag{3.48}$$

Aside from showing another interesting difference between the Chern-Simons term and the Maxwell term, we can use this inverse to couple the Chern-Simons theory to an external electromagnetic gauge potential $\mathcal{A}_\mu$. We will set $e = \hbar = 1$ except as noted. We do not include the Maxwell term to give this field dynamics, but rather view it as an imposed field *beyond the magnetic field producing the phase*. For example,

we could use this additional field to add an electrical field, and we should find a Hall response. Let's try this:

$$\mathcal{L} = \frac{k}{4\pi}\epsilon^{\mu\nu\lambda}a_\mu\partial_\nu a_\lambda - \frac{1}{2\pi}\epsilon^{\mu\nu\lambda}A_\mu\partial_\nu a_\lambda = \frac{k}{4\pi}\epsilon^{\mu\nu\lambda}a_\mu\partial_\nu a_\lambda - \frac{1}{2\pi}\epsilon^{\mu\nu\lambda}a_\mu\partial_\nu A_\lambda, \quad (3.49)$$

where in the second step we have dropped a boundary term and used the antisymmetry property of the ϵ tensor. Note that to obtain the second term we have just rewritten $A_\mu J^\mu$ using (3.40.

Now we can integrate out a_μ using equation (3.48) above, recalling $\gamma = k/(8\pi)$, and obtain

$$\mathcal{L}_{\text{eff}} = \frac{\pi}{k}J_\mu\epsilon^{\mu\nu\lambda}\partial_\nu\frac{1}{\partial^2}J_\lambda = \frac{1}{4\pi k}\epsilon^{\mu\alpha\beta}\partial_\alpha A_\beta\epsilon^{\mu\nu\lambda}\partial_\nu\frac{1}{\partial^2}\epsilon^{\lambda\gamma\delta}\partial_\gamma A_\delta. \quad (3.50)$$

where in the second step we have used the rewritten Lagrangian in (3.49) to identify $J^\mu = \frac{1}{2\pi}\epsilon^{\mu\nu\lambda}\partial_\nu A_\lambda$. As above, the nonzero possibilities are $\alpha = \nu$ and $\beta = \lambda$ (+1) or vice versa (-1), and also $\gamma = \mu$ and $\delta = \nu$ (+1) or vice versa (-1). Working through these, one is left with the $\gamma = \nu$ and $\delta = \mu$ terms,

$$\mathcal{L}_{\text{eff}} = \frac{1}{4\pi k}\epsilon^{\mu\nu\lambda}A_\mu\partial_\nu A_\lambda. \quad (3.51)$$

This is the *electromagnetic* Chern-Simons term. The electromagnetic current is obtained by varying A:

$$J^\mu = -\frac{\delta\mathcal{L}_{\text{eff}}}{\delta A_\mu} = \frac{1}{2\pi k}\epsilon^{\mu\nu\lambda}\partial_\nu A_\lambda. \quad (3.52)$$

where the factor of 2 is obtained because the variation can act on either A.

We can see immediately that this predicts a Hall effect: in response to an electrical field along x, we obtain a current along y. What about the factor $1/(2\pi)$? That is here just so that the response, once we restore factors of e and $\hbar$, is

$$\sigma_{xy} = \frac{e^2}{(2\pi)k\hbar} = \frac{1}{k}\frac{e^2}{h}. \quad (3.53)$$

Here we get a clue about the physical significance of k. Another clue is to consider the electromagnetic charge J^0 induced by a change in the magnetic field δB (i.e., an additional field beyond the one producing the FQHE):

$$J^0 = \delta n = \frac{1}{2\pi k}\delta B. \quad (3.54)$$

where we have written $J^0 = \delta n$ to indicate that this electromagnetic density describes the change in electron density from the ground state without the additional field. For the IQHE, a change of one flux quantum corresponds to one additional electron, while we can see that the $k = 3$ Chern-Simons theory predicts a change in density $e/3$, consistent with the quasihole and quasiparticle excitations.

To summarize what we have learned so far, we now see that Chern-Simons theory predicts a connection between the Hall quantum, the statistics of quasiparticles in the theory (from the previous section), and the effective density induced by a local change in the magnetic field. Here "quasiparticles", which we will discuss later, means whatever particle couples to the Chern-Simons theory as in the preceding section, which need not be an electron.

3.3.3 Chern-Simons theory III: topological aspects and gapless edge excitations

One obvious respect in which the Chern-Simons theory is topological is that, because ϵ rather than the metric tensor g was used to raise the indices, there is no dependence on the metric. In Zee's language, it describes a world without rulers or clocks. Since the stress-energy tensor in a relativistic theory is determined by varying the Lagrangian with respect to the metric, the stress-energy tensor is identically zero.

How can a theory be interesting if all its states have zero energy, as in the pure Chern-Simons theory? Well, one interesting fact is that the number of zero-energy states is dependent on the manifold where the theory is defined. We will not try to compute this in general but will solve the theory for the case of the torus. It is quite surprising that we can solve this 2+1-dimensional field theory exactly; the key will be that there are very few physical degrees of freedom once the $U(1)$ gauge invariance is taken into account.

We wish to solve the pure Chern-Simons theory with action

$$\mathcal{L}_{CS} = \frac{k}{4\pi}\epsilon^{\mu\nu\lambda}a_\mu\partial_\nu a_\lambda \tag{3.55}$$

on the manifold $\mathbb{R}$(time) $\times\, T^2$(space). The gauge invariance is under $a_\mu \to a_\mu + \partial_\mu\chi$, χ an arbitrary scalar function. Given an arbitrary configuration of the gauge field a_μ, we first fix $a_0 = 0$ by the gauge transformation $a_\mu \to a_\mu + \partial_\mu\chi$ with $\chi = -\int a_0\, dt$. The Lagrangian is then

$$\mathcal{L} = -\frac{k}{4\pi}\epsilon^{ij}a_i\dot{a}_j, \tag{3.56}$$

where $i, j = 1, 2$. The equation of motion from varying the original Lagrangian with respect to a_0 now gives a constraint

$$\epsilon_{ij}\partial_i a_j = 0. \tag{3.57}$$

There is still some gauge invariance remaining in a_1, a_2: we can add a purely spatially dependent χ, so that a_0 remains 0, to make $\partial_i a_i = 0$ (exercise). Then $(a_i(t), a_j(t))$ have zero spatial derivatives and hence are purely functions of time. The Lagrangian (3.56) is now just the minimal coupling of a particle moving in a position-dependent vector potential; thinking of (a_1, a_2) as the coordinates of a particle moving in the plane, and noting that a constant magnetic field can be described by the vector potential $(By/2, -Bx/2) = (Ba_2/2, -Ba_1/2)$, we see that this is the interaction term of a particle in a constant magnetic field.

So far, using gauge invariance we can reduce the degrees of freedom from a 2+1-dimensional field theory to the path integral for the quantum mechanics of a particle moving in two dimensions. There is one last bit of gauge invariance we need to use. This will reduce the space on which our particle moves, which so far is $\mathbb{R}^2$ because the gauge fields are noncompact, to the torus T^2 on which the theory is defined. We consider a gauge transformation of the form $a_j \to a_j - iu^{-1}\partial_j u$, where u is purely a function of space. Note that if we can write $u = \exp(i\theta)$, this becomes a conventional gauge transformation $a_j \to a_j + \partial_j\theta$. This gauge transformation will not break the previous two gauge constraints if $\nabla^2\theta = 0$.

However, the periodicity of the torus means that we might not be able to define θ periodically, even if u is defined globally and the gauge transformation is indeed periodic. Taking the torus to be $L_1 \times L_2$, the following θ has zero Laplacian everywhere and gives rise to a periodic u and hence a periodic gauge transformation, even if θ is not itself periodic:

$$\theta = \frac{2\pi n_1 x}{L_1} + \frac{2\pi n_2 y}{L_2}. \tag{3.58}$$

The effect of this gauge transformation is that we can shift the particle's trajectory by an arbitrary constant integer multiple of L_1 in the x direction and L_2 in the y direction. To make the torus equivalent to the unit torus, we can rescale $a_i(t) = (2\pi/L_i)q_i(t)$. So finally we have shown

$$S = \int d^2x\, dt\, \frac{k}{4\pi}\epsilon^{\mu\nu\lambda}a_\mu\partial_\nu a_\lambda = -\frac{kL_1L_2}{4\pi}\int dt\, \frac{(2\pi)^2}{L_1L_2}\epsilon^{ij}q_i\dot{q}_j. \tag{3.59}$$

Here one L_1L_2 factor is from the spatial integrals and one is from the change of variable from a_i to q_i. We still haven't done anything quantum-mechanical to solve the path integral. However, we can temporarily add a term $m\dot{q}_i^2/2$ to the Lagrangian and recognize it as the path integral for a particle moving on the torus in a constant magnetic field. The gauge potential is $A_i = k\pi\epsilon_{ij}q_j$, which corresponds to a magnetic field $B = 2\pi k$ (this factor of 2 always appears in the rotational gauge). This is in our theorist's units with $\hbar = e = 1$; it means that there are a total of k flux quanta through the torus.

The limit we care about for pure CS theory is $m \to 0$, which takes all states not in the lowest Landau level to infinite energy. This makes sense because in a topological theory there can be no energy scale; the states either have some constant energy (the lowest Landau level here), which can be taken to zero, or infinite energy (the other Landau levels here). A quick calculation shows that there are exactly k states in the lowest Landau level on the torus pierced by k flux quanta; note that the "shift" of 1 extra level on the sphere is absent. For example, the lowest Landau level with one flux quantum through the *sphere* corresponds to the coherent-state path integral for a $s = 1/2$ particle (see problem sets), with 2 degenerate states.

The conclusion is that the parameter k also controls the ground-state degeneracy on the torus. An argument (X.G. Wen and Q. Niu, Phys. Rev. B41, 9377 (1990)) (regrettably direct calculation seems to be more difficult) shows that the general degeneracy of the pure Abelian CS theory on a 2-manifold of genus g is k^g. So for a topological theory, the physical content of the model is determined not just by explicit parameters in the action, such as k, but also by the topology of the manifold where the theory is defined. In this sense topological theories are sensitive to global or "long-ranged" properties, even though the theory is massive/gapped. (Of course, in the pure CS theory there is no notion of length so the distinction between local and global doesn't mean much, but adding a Maxwell term or something like that would not modify the long-distance properties; it would just mean that the other Landau levels are no longer at infinite energy.)

3.3.4 Bulk-edge correspondence

We noted above that the Chern-Simons term has different gauge-invariance properties from the Maxwell term: in particular, in a system with a boundary, it is not gauge-invariant by itself because the boundary term we found above need not vanish. Our last goal in this section is to see that this gauge invariance leads to the free massless chiral boson theory at the edge,

$$S_{\text{edge}} = \frac{k}{4\pi} \int dt\, dx\, (\partial_t + v\partial_x)\phi\partial_x\phi. \tag{3.60}$$

Here k is exactly the same integer coefficient as in the bulk CS theory, while v is a nonuniversal velocity that depends on the confining potential and other details. Note that the kinetic term here is "topological" in the sense that it does not contribute to the Hamiltonian, because it is first-order in time. The second term is not topological and hence shouldn't be directly obtainable from the bulk theory.

The theory of the bulk and boundary is certainly invariant under "restricted" gauge transformations that vanish at the boundary: $a_\mu \to a_\mu + \partial_\mu\chi$ with $\chi = 0$ on the boundary. From (3.36) above, the boundary term vanishes if $\chi = 0$ there. This constraint means that degrees of freedom that were previously gauge degrees of freedom now become dynamical degrees of freedom. We will revisit this idea later.

To start, choose the gauge condition $a_0 = 0$ as in the previous section and again use the equation of motion for a_0 as a constraint. [3] Then $\epsilon^{ij}a_j = 0$ and we can write $a_i = \partial_i\phi$. Substituting this into the bulk Chern-Simons Lagrangian

$$\begin{aligned} S &= -\frac{k}{4\pi}\int \epsilon^{ij} a_i\partial_0 a_j\, d^2x\, dt = -\frac{k}{4\pi}\int (\partial_x\phi\partial_0\partial_y\phi - \partial_y\phi\partial_0\partial_x\phi)\, d^2x\, dt \\ &= -\frac{k}{4\pi}\int (\partial_x(\phi\partial_0\partial_y\phi) - \partial_y(\phi\partial_0\partial_x\phi))\, d^2x\, dt \\ &= -\frac{k}{4\pi}\int (\nabla\times\mathbf{v})_z\, d^2x\, dt = -\frac{k}{4\pi}\int \mathbf{v}\cdot d\mathbf{l}\, dt, \end{aligned} \tag{3.61}$$

where $\mathbf{v}$ is the vector field

$$\mathbf{v} = (\phi\partial_0\partial_x\phi, \phi\partial_0\partial_y\phi). \tag{3.62}$$

(You might wonder why this doesn't let us transform the action simply to zero in the case of the torus studied in the previous section. The reason is that using Stokes's theorem in the second line, we have assumed the disk topology–since the torus has nontrivial topology, we are not allowed to use Stokes's theorem to obtain zero, cf. "Preliminaries" lecture notes.) So at the boundary, which we will assume to run along x for compactness, the resulting action is, after an integration by parts,

$$S_{\text{edge}} = \frac{k}{4\pi}\int \partial_t\phi\, \partial_x\phi\, dx\, dt. \tag{3.63}$$

We're almost done–this predicts a "topological" edge theory determined by the bulk physics; this edge theory is topological in that the Hamiltonian is zero. However,

[3] Here and before we are assuming that the Jacobians from our gauge-fixings and changes of variables are trivial. That this is the case is argued in S. Elitzur et al., Nuclear Physics B **326**, 108 (1989). Another nice discussion in this paper is how, for the non-Abelian case, the bulk can be understood as providing the Wess-Zumino term that keeps the edge theory gapless.

in order to obtain an accurate physical description we need to include non-universal, non-topological physics arising from the details of how the Hall droplet is confined. One approach to this is to start from a hydrodynamical theory of the edge and then recognize one term in that theory as $S_{\rm edge}$ above. The other term in that theory is a nonuniversal velocity term, and the combined action is

$$S_{\rm edge} = \frac{k}{4\pi} \int (\partial_t \phi - v \partial_x \phi)\, \partial_x \phi \, dx \, dt. \tag{3.64}$$

Here the nonuniversal parameter v clearly has units of a velocity, and in the correlation functions of the theory discussed below indeed appears as a velocity. The Hamiltonian density is

$$\mathcal{H} = \frac{kv}{4\pi} (\partial_x \phi)^2 \tag{3.65}$$

Note that for the Hamiltonian to be positive definite, the product kv needs to be positive: in other words, the sign of the velocity is determined by the bulk parameter k even thought the magnitude is not, and the edge is indeed chiral. (The density at the edge is found from the hydrodynamical argument to be proportional to $\partial_x \phi/(2\pi)$, so the above interaction term corresponds to a short-ranged density-density interaction; as usual, we will neglect the differences that arise if the long-ranged Coulomb interaction is retained instead.)

3.3.5 Chern-Simons theory IV: connecting edge theory to observables

We give a quick overview of how the above theory leads to detailed predictions of several edge properties. The general approach to treating one-dimensional electronic systems via free boson theories is known as "bosonization", and is the subject of several books. [4]. While we will not calculate the main results in detail, it turns out that there is a close similarity between the 1-dimensional free (chiral or nonchiral) boson Lagrangian and the theory of the algebraic phase of the XY model studied previously.

The reason such a connection exists is simple: the Euclidean version of the nonchiral version of the above free boson theory is just the 2D Gaussian theory. However, we know from the study of the XY model that subtleties such as the Berezinskii-Kosterlitz-Thouless transition arise when the variable appearing in the Gaussian theory is taken to be periodic, as when it describes an angular variable in that model. One of the surprising results we found was a power-law phase with continuously variable exponents: the correlations of spin operators $S_x + iS_y = \exp(i\theta)$ go as a power-law with the coefficient depending on the prefactor of the Gaussian.

The connection between the edge theory above and physical quantities is that the electron correlation function is represented in the bosonized theory as $e^{ik\phi}$: effectively ϕ describes a single quasiparticle and k quasiparticles make up the electron. The electron propagator in momentum space is likewise here found to have an exponent that depends on k: there is a factor of k^2 from the k's in the electron operator, and a

[4] For example, M. Stone, *Bosonization*, World Scientific

factor of k^{-1} from the quasiparticle propagator since k appears as a coefficient in the Lagrangian. The result is

$$G(q,\omega) \propto \frac{(vq+\omega)^{k-1}}{vq-\omega}. \tag{3.66}$$

This describes an electron density of states $N(\omega) \propto |\omega|^{k-1}$, and this exponent can be measured in tunneling exponents: $dI/dV \propto V^{k-1}$. As a sanity check, the $k=1$ case describes a constant density of states and the predicted conduction is Ohmic: $I \propto V$.

Experimental agreement is reasonable but hardly perfect; at $\nu = 1/3$ the observed tunneling exponent $I \propto V^{\alpha}$ observes $\alpha \approx 2.7$, which is far from the Ohmic value ($\alpha = 1$) but reasonably close to the predicted value $\alpha = 3$. The tunneling exponent also does not appear to be perfectly constant when one is on a Hall plateau, as the theory would predict. Other measurements include "noise" measurements that attempt to see the quasiparticle charge directly, and in recent years interferometry measurements that try to check more subtle aspects of the theory.

In closing we comment briefly on the generalization of the above Chern-Simons and edge theories to more complicated (but still Abelian) quantum Hall states. These states, as suggested by the hierarchy picture, have multiple types of "particles", and two particles can have nontrivial statistics whether or not they belong to the same species. These statistics are defined by a universal integer "K matrix" that can be taken as a fundamental aspect of the topological order in the state. (Information must also be provided about the allowed quasiparticle types.) The resulting CS theory is

$$\mathcal{L} = \frac{1}{4\pi} K^{IJ} a^I_\mu \partial_\nu a^J_\lambda \tag{3.67}$$

This effective theory works for all but a few proposed quantum Hall states. I believe Prof. Stern will discuss these exotic "non-Abelian" quantum Hall states elsewhere in this volume.

Topological invariants in 2D with time-reversal invariance

.0.1 An interlude: Wess-Zumino terms in one-dimensional nonlinear σ-models

A mathematical strategy similar to what we will need for the QSHE was developed by Wess and Zumino in the context of 1+1-dimensional field theory. The free boson is described by the action

$$S_0 = -\frac{K}{2}\int_{\mathbb{R}^2} (\nabla\phi)^2, \tag{.68}$$

which for a compact boson field ϕ is the nonlinear sigma model into the circle S^1, which is the manifold of the Lie group $U(1)$. The direct generalization of this to a more complicated Lie group such as $SU(N)$ is written as

$$S_0 = -\frac{k}{8\pi}\int_{S^2} \mathcal{K}(g^{-1}\partial^\mu g, g^{-1}\partial_\mu g), \tag{.69}$$

where we have compactified the plane to the sphere, changed the prefactor, and written the interaction in terms of the "Killing form" $\mathcal{K}$ on the Lie algebra associated with g.

(This Killing form is a symmetric bilinear form that, in the $U(1)$ case above, is just the identity matrix.) Unfortunately this action behaves quite differently from the $U(1)$ case: it does not describe a critical theory (in particle physics language, it develops a mass).

To fix this problem, Wess and Zumino wrote a term

$$S_{WZ} = -\frac{2\pi k}{48\pi^2}\int_{B^3}\epsilon_{\mu\nu\lambda}\mathcal{K}\left(g^{-1}\partial_\mu g, \left[g^{-1}\partial_\nu g, g^{-1}\partial_\lambda g\right]\right) \tag{.70}$$

that is quite remarkable: even writing this term depends on being able to take an original configuration of g on the sphere S^2 and extend it in to the sphere's interior B^3. (We will not show here that this term accomplishes the desired purpose, just that it is topologically well-defined.) At least one contraction into the ball exists because $\pi_2(G) = 0$. Different contractions exist because $\pi_3(G) = \mathbb{Z}$, and the coefficient of the second term is chosen so that, if k (the "level" of the resulting Wess-Zumino-Witten theory) is an integer, the different topological classes differ by a multiple of $2\pi i$ in the action, so that the path integral is independent of what contraction is chosen. The reason that $\pi_3(G)$ is relevant here is that two different contractions into the interior B^3 can be joined together at their common boundary to form a 3-sphere, in the same way as two disks with the same boundary can be joined together to form the top and bottom hemispheres of a 2-sphere.

.0.2 Topological invariants in time-reversal-invariant Fermi systems

The main subtlety in finding a topological invariant for time-reversal-invariant band structures will be in keeping track of the time-reversal requirements. We introduce $\mathcal{Q}$ as the space of time-reversal-invariant Bloch Hamiltonians. This is a subset of the space of Bloch Hamiltonians with at most pairwise degeneracies (the generalization of the nondegenerate case we described above; we need to allow pairwise degeneracies because bands come in Kramers-degenerate pairs). In general, a $\mathcal{T}$-invariant system need not have Bloch Hamiltonians in $\mathcal{Q}$ except at the four special points where $k = -k$. The homotopy groups of $\mathcal{Q}$ follow from similar methods to those used above: $\pi_1 = \pi_2 = \pi_3 = 0$, $\pi_4 = \mathbb{Z}$. $\mathcal{T}$-invariance requires an even number of bands $2n$, so $\mathcal{Q}$ consists of $2n \times 2n$ Hermitian matrices for which H commutes with Θ, the representation of $\mathcal{T}$ in the Bloch Hilbert space:

$$\Theta H(k)\Theta^{-1} = H(-k). \tag{.71}$$

Our goal in this section is to give a geometric derivation of a formula, first obtained by Fu and Kane via a different approach, for the $\mathbb{Z}_2$ topological invariant in terms of the Berry phase of Bloch functions:

$$D = \frac{1}{2\pi}\left[\oint_{\partial(EBZ)} d\boldsymbol{k}\cdot\boldsymbol{\mathcal{A}} - \int_{EBZ} d^2k\,\mathcal{F}\right] \mod 2. \tag{.72}$$

The notation EBZ stands for Effective Brillouin Zone, [**?**] 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. .2(b). While $\mathcal{F}$ is gauge-invariant, $\boldsymbol{\mathcal{A}}$ is not, and different (time-reversal-invariant) gauges, in a sense made precise below, can change the boundary integral by an even amount. The formula (.72) was not the first definition of the two-dimensional $\mathbb{Z}_2$ invariant, as the original Kane-Mele paper gave a definition based on counting of zeros of the "Pfaffian bundle" of wavefunctions. However, (.72) is both easier to connect to the IQHE and easier to implement numerically.

The way to understand this integral is as follows. Since the EBZ has boundaries, unlike the torus, there is no obvious way to define Chern integers for it; put another way, the $\mathcal{F}$ integral above is not guaranteed to be an integer. However, given a map from the EBZ to Bloch Hamiltonians, we can imitate the Wess-Zumino approach above and consider "contracting" or "extending" the map to be one defined on the sphere (Fig. .3), by finding a smooth way to take all elements on the boundary to some constant element $\mathcal{Q}_0 \in \mathcal{Q}$. The geometric interpretation of the line integrals of $\mathcal{A}$ in (.72) is that these are the integrals of $\mathcal{F}$ over the boundaries, and the requirement on the gauge used to define the two $\mathcal{A}$ integrals is that each extends smoothly in the associated cap. The condition on the cap is that each vertical slice satisfy the same time-reversal invariance condition as an EBZ boundary; this means that a cap can alternately be viewed as a way to smoothly deform the boundary to a constant, while maintaining the time-reversal condition at each step.

The two mathematical steps, as in the Wess-Zumino term, are showing that such contractions always exist and that the invariant D in (.72) is invariant of which contraction we choose. The first is rather straightforward and follows from $\pi_1(\mathcal{H}) = \pi_1(\mathcal{Q}) = 0$. The second step is more subtle and gives an understanding of why only a $\mathbb{Z}_2$ invariant or "Chern parity" survives, rather than an integer-valued invariant as the IQHE. We can combine two different contractions of the same boundary into a sphere, and the Chern number of each band pair on this sphere gives the difference between the Chern numbers of the band pair obtained using the two contractions (Fig. .3).

The next step is to show that the Chern number of any band pair on the sphere is even. To accomplish this, we note that Chern number is a homotopy invariant and that it is possible to deform the Bloch Hamiltonians on the sphere so that the equator is the constant element $\mathcal{Q}_0$ (here the equator came from the time-reversal-invariant elements at the top and bottom of each allowed boundary circle.) The possibility of deforming the equator follows from $\pi_1(\mathcal{Q}) = 0$, and the equivalence of different ways of deforming the equator follows from $\pi_2(\mathcal{Q}) = 0$. Then the sphere can be separated into two spheres, related by time-reversal, and the Chern numbers of the two spheres are equal so that the total Chern number is zero.

The above argument establishes that the two values of the $\mathbb{Z}_2$ invariant are related to even or odd Chern number of a band pair on half the Brillouin zone. Note that the lack of an integer-valued invariant means, for example, that we can smoothly go from an S_z-conserved model with $\nu = 1$ for spin $\uparrow$, $\nu = -1$ for spin $\downarrow$ to a model with $\nu = \pm 3$ by breaking S_z conservation in between. This can be viewed as justification for the physical argument given above in terms of edge states annihilating in pairs, once we define the $\mathbb{Z}_2$ invariant for disordered systems in the following section.

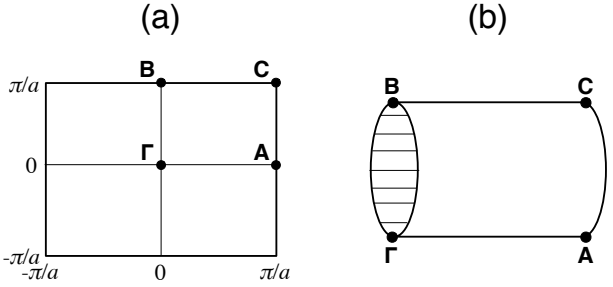

Fig. .2 (a) A two-dimensional Brillouin zone; note that any such Brillouin zone, including that for graphene, can be smoothly deformed to a torus. The labeled points are time-reversal-invariant momenta. (b) The effective Brillouin zone (EBZ). The horizontal lines on the boundary circles ∂(EBZ) connect time-reversal-conjugate points, where the Hamiltonians are related by time reversal and so cannot be specified independently.

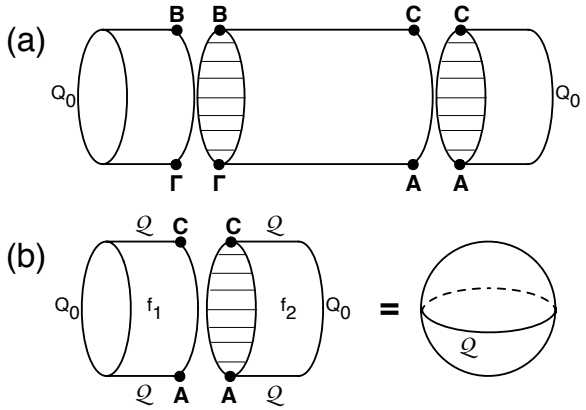

Fig. .3 (a) Contracting the extended Brillouin zone to a sphere. (b) Two contractions can be combined to give a mapping from the sphere, but this sphere has a special property: points in the northern hemisphere are conjugate under $\mathcal{T}$ to those in the southern, in such a way that overall every band pair's Chern number must be even.

.0.3 Pumping interpretation of $\mathbb{Z}_2$ invariant

We expect that, as for the IQHE, it should be possible to reinterpret the $\mathbb{Z}_2$ invariant as an invariant that describes the response of a finite toroidal system to some perturbation. In the IQHE, the response is the amount of charge that is pumped around one circle of the torus as a 2π flux (i.e., a flux hc/e) is pumped adiabatically through the other circle. [5] Here, the response will again be a pumped charge, but the cyclic process that pumps the chage is more subtle.

[5] A previous pumping definition that involves a π-flux but considers pumping of "$\mathbb{Z}_2$" from one boundary to another of a large cylinder was given by Fu and Kane.

Instead of inserting a 2π flux through a circle of the toroidal system, we insert a π flux, adiabatically; this is consistent with the part of D in (.72) that is obtained by integration over half the Brillouin zone. However, while a π flux is compatible with T-invariance, it is physically distinct from zero flux, and hence this process is not a closed cycle. We need to find some way to return the system to its initial conditions. We allow this return process to be anything that does not close the gap, but require that the Hamiltonians in the return process *not* break time-reversal. Since the forward process, insertion of a π flux, definitely breaks time-reversal, this means that the whole closed cycle is a nontrivial loop in Hamiltonian space. The $\mathbb{Z}_2$ invariant then describes whether the charge pumped by this closed cycle through the other circle of the torus is an odd or even multiple of the electron charge; while the precise charge pumped depends on how the cycle is closed, the parity of the pumped charge (i.e., whether it is odd or even) does not.

This time-reversal-invariant closure is one way to understand the physical origin of the $\mathcal{A}$ integrals in (.72), although here, by requiring a closed cycle, we have effectively closed the EBZ to a torus rather than a sphere. One weakness of the above pumping definition, compared to the IQHE, is that obtaining the $\mathbb{Z}_2$ invariant depends on Fermi statistics, so that the above pumping definition cannot be directly applied to the many-body wavefunction as in the IQHE case. We will solve this problem later for the three-dimensional topological insulator by giving a pumping-like definition that can be applied to the many-particle wavefunction.

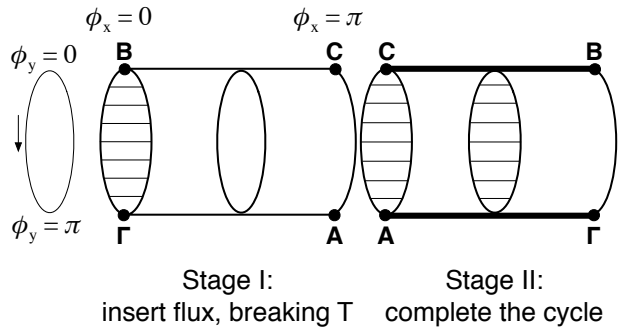

Fig. .4 Graphical representation of charge pumping cycle for Chern parities. The first stage takes place as the flux ϕ_x increases adiabatically from 0 to π. In the second stage the Hamiltonian at $(\phi_x = \pi, \phi_y)$ is adiabatically transported through the space of Hamiltonians to return to the 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.