Notes for MIT minicourse on topological phases

Joel E. Moore

(Dated: January 17, 2011)

Outline of what was covered in lectures:

1A. Physical background of IQHE and topological insulators (slides). Crash course on homotopy and cohomology. Examples: equivalence of Chern number in a 2D two-band model to $\pi_2(S^2)$ (in notes on page). Sketch of Haldane's theory for gap in integer-spin 1D Heisenberg antiferromagnets.

1B. Meaning of Berry phase of Bloch wavefunctions: semiclassical equations of motion.

2. FQHE physical background (in textbooks, so not really covered in slides or on board). Two FQHE pictures: Laughlin state via Haldane pseudopotentials. Composite fermions. Meaning of fractional statistics.

3. Chern-Simons internal theory and flux attachment.

A. Cohomology example: invariant integrals along paths in two dimensions and 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:

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

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.

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

$$\partial_x g = \lim_{\Delta x \to 0} \frac{g(x + \Delta x, y) - g(x, y)}{\Delta x}. \tag{2}$$

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

$$f = dg = \partial_x g\, dx + \partial_y g\, dy. \tag{3}$$

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 $x_1 = x, x_2 = y, f_1 = p, f_2 = q$)

$$df = \sum_j \frac{\partial f_i}{\partial x_j}\, dx_j \wedge dx_i. \tag{4}$$

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,

$$df = (\partial_x f_y - \partial_y f_x) dx \wedge dy, \tag{5}$$

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.

B. 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

$$f = (p, q) = \left(-\frac{y}{x^2 + y^2}, \frac{x}{x^2 + y^2} \right) = \frac{-y dx + x dy}{x^2 + y^2}, \tag{6}$$

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:

$$df = \left(\partial_x \frac{x}{x^2 + y^2} \right) dx \wedge dy + \left(\partial_y \frac{-y}{x^2 + y^2} \right) dy \wedge dx = \frac{2 - 2}{x^2 + y^2} dx \wedge dy = 0. \tag{7}$$

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. 6 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),

$$H^1(M) = \frac{\text{closed 1-forms on } M}{\text{exact 1-forms on } M} = \frac{Z^1(M)}{B^1(M)}. \tag{8}$$

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. 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. 6, 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

$$c = \frac{\int_0^{2\pi} f'}{\int_0^{2\pi} f}. \tag{9}$$

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

$$g(\theta) = \int_0^{\theta} (f' - cf). \tag{10}$$

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

$$\chi(M) = \sum_i (-1)^i \dim H^i(M) = \sum_i (-1)^i \dim \frac{Z^i(M)}{B_i(M)}. \tag{11}$$

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 that comes in later. 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:

$$\begin{array}{l} \longleftarrow C_r \xleftarrow[\partial_{r+1}]{} C_{r+1} \xleftarrow[\partial_{r+2}]{} C_{r+2} \longleftarrow \\ \longrightarrow \Omega_r \xrightarrow[d_{r+1}]{} \Omega_{r+1} \xrightarrow[d_{r+2}]{} \Omega_{r+2} \longrightarrow . \end{array} \tag{12}$$

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

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

$$(f, \partial c) = (df, c). \tag{13}$$

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.

C. 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

$$f(x,0)=\gamma_1(x), \quad f(x,1)=\gamma_2(x). \tag{14}$$

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 totally independent of the cohomology groups: for example, if $\pi_1(M)$ is trivial, then so is the first de Rham group. 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.

Homotopy was the original application of topology in condensed matter physics: it can be used to classify topological defects such as vortices in broken-symmetry phases. The review article of Mermin is a good reference for this topic. Briefly, a topological defect that can be enclosed by a d-dimensional sphere is classified by the homotopy group $\pi_d(M)$, where M is the order parameter manifold (see Appendix). $M=G/H$, where G is the high-temperature symmetry group (i.e., the symmetry group of the disordered phase), H is the residual symmetry of the ordered phase, and division means forming cosets (identifying two states that differ by a symmetry in H).

I. 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 subsection 2. Before that, we give some background for the original quantum Hall discovery that triggered a flood of developments continuing to the present day.

A. Physical background of the IQHE

(Here we follow standard treatments, e.g., the introductions to the books by Prange and Girvin, or Das Sarma and Pinczuk, so I will not write new notes.)

B. 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

$$\psi_{\mathbf{k}}(\mathbf{r})=e^{i\mathbf{k}\cdot\mathbf{r}}u_{\mathbf{k}}(\mathbf{r}), \tag{15}$$

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

$$(T\psi)(x) = \lambda\psi(x). \tag{16}$$

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

$$\psi_k(x+a) = e^{ika}\psi_k(x). \tag{17}$$

The last step is to define $u_k(x) = \psi_k(x)e^{-ikx}$; then (17) 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"' $k \in [-\pi/a, \pi/a)$:

$$\gamma = \oint_{-\pi/a}^{\pi/a} \langle u_k | i\partial_k | u_k \rangle dk. \tag{18}$$

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, $|u_k\rangle \to e^{-i\phi(k)}|u_k\rangle$, with $\phi(\pi/a) = \phi(-\pi/a) + 2\pi n, n \in \mathbb{Z}$? Under this transformation, the integral shifts as

$$\gamma \to \gamma + \oint_{-\pi/a}^{\pi/a} (\partial_k \phi)\, dk = \gamma + 2\pi n. \tag{19}$$

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.

---

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

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

$$n_j = \frac{i}{2\pi} \int dk_x \, dk_y \, \left(\langle \partial_{k_x} u_j | \partial_{k_y} u_j \rangle - \langle \partial_{k_y} u_j | \partial_{k_x} u_j \rangle \right) \tag{20}$$

to the total Hall conductance:

$$\sigma_{xy} = \frac{e^2}{h} \sum_j n_j. \tag{21}$$

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.

C. 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:

$$\Delta P = \int_0^1 d\lambda \frac{dP}{d\lambda} = \int_{t_0}^{t_1} dt \, \frac{dP}{d\lambda} \frac{d\lambda}{dt} = \int_{t_0}^{t_1} j(t) \, dt. \tag{22}$$

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

$$H(q, \lambda) = \frac{1}{2m}(p + \hbar q)^2 + V^{(\lambda)}(r). \tag{23}$$

The current operator is

$$j(q) = ev(q) = \frac{ie}{\hbar}[H(q, \lambda), r] = \frac{e}{m}(p + \hbar q) = \frac{e}{\hbar} \partial_q H(q, \lambda). \tag{24}$$

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

$$|\psi(t)\rangle = \exp\left(-(i/\hbar) \int^t E_0(t') \, dt' \right) \left[|\psi_0(t)\rangle + i\hbar \sum_{j \neq 0} |\psi_j(t)\rangle (E_j - E_0)^{-1} \langle \psi_j(t) | \dot{\psi}_0(t) \rangle \right]. \tag{25}$$

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

$$\frac{dP}{d\lambda} = 2\hbar e \, \mathrm{Im} \sum_{v,c} \int \frac{dq}{2\pi} \frac{\langle u_v(q) | v(q) | u_c(q) \rangle \langle u_c(q) | \partial_\lambda u_v(q) \rangle}{E_c(q) - E_v(q)}. \tag{26}$$

For example, we wrote

$$\langle \psi_j(t)|\dot{\psi}_0(t)\rangle = \sum_{v,c}\langle u_c|\partial_\lambda u_v\rangle \frac{d\lambda}{dt}. \tag{27}$$

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 $H' = dq\,\partial_q H$:

$$|\partial_q u_j(q)\rangle = \sum_{j\neq j'} |u_{j'}(q)\rangle \frac{\langle u_{j'}(q)|\partial_q H(q,\lambda)|u_j(q)\rangle}{E_j(q) - E_{j'}(q)}. \tag{28}$$

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

$$\frac{dP}{d\lambda} = 2\hbar e\,\mathrm{Im}\sum_c \int \frac{dq}{2\pi}\frac{\langle u_v(q)|v(q)|u_c(q)\rangle\langle u_c(q)|\partial_\lambda u_v(q)\rangle}{E_c(q) - E_v(q)} = 2e\,\mathrm{Im}\int \frac{dq}{2\pi}\langle \partial_q u_v(q)|\partial_\lambda u_v(q)\rangle. \tag{29}$$

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

$$\Delta P = 2e\,\mathrm{Im}\int_0^1 d\lambda \int \frac{dq}{2\pi}\langle \partial_q u_v(q)|\partial_\lambda u_v(q)\rangle. \tag{30}$$

The last expression is in two dimensions and involves the same type of integrand (a Berry flux) as in the 2D TKNN formula (20). 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,

$$P = ie\int \frac{dq}{2\pi}\langle u_v(q)|\partial_q u_v(q)\rangle, \tag{31}$$

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.

D. Interactions and disorder: the flux trick

One might worry whether the TKNN integer defined in equation (20) 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:

$$\frac{dP_y}{dt} = j_y = \frac{dP_y}{d\Phi_x}\frac{d\Phi_x}{dt} = \frac{dP_y}{d\Phi_x}(cE_x L_x). \tag{32}$$

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)

$$P_y(\Phi_x) = \frac{ie}{2\pi} \int d\Phi_y \, \langle u | \partial_{\Phi_y} u \rangle \tag{33}$$

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:

$$\langle j_y \rangle = \langle \frac{dP_y}{d\Phi_x} \rangle (cE_x L_x) \to \frac{\Delta P_y}{\Delta \Phi_x} (cE_x L_x), \tag{34}$$

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

$$\langle j_y \rangle = \frac{e}{hc} \frac{ne}{L_x} (cE_x L_x) = \frac{ne^2}{h} E_x. \tag{35}$$

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 (20), now in terms of fluxes.

E. 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

$$n_j = \frac{i}{2\pi} \int d^2k \, \mathrm{Tr} \, [dP_j \wedge P_j \, dP_j], \tag{36}$$

where $\wedge$ is the wedge product and $dP_j = \partial_{k_x} P_j \, dx + \partial_{k_y} P_j \, dy$ is a differential form where the coefficients are operators. (Note that the wedge product in the above formula acts only on dx and dy.) It is a straightforward exercise to verify that this reproduces the TKNN definition (20).

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

$$H(k_x, k_y) = a_0(k_x, k_y)\mathbf{1} + a_1(k_x, k_y)\sigma_x + a_2(k_x, k_y)\sigma_y + a_3(k_x, k_y)\sigma_z. \tag{37}$$

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

$$(a_1, a_2, a_3) \to \frac{(a_1, a_2, a_3)}{\sqrt{{a_1}^2 + {a_2}^2 + {a_3}^2}} \tag{38}$$

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.

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; we return to the Hopf invariant below when we discuss Chern-Simons terms in three dimensions.

Now we consider the case of a nondegenerate two-dimensional band structure with multiple bands. 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:

$$H(k) = U(k)D(k)U^{-1}(k). \tag{39}$$

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.

F. 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 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, and others, 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

---

[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.

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

$$J^i_j = \alpha_s \epsilon_{ijk} E_k, \tag{40}$$

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. 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:

$$H_0 = t \sum_{\langle ij \rangle \sigma} c^\dagger_{i\sigma} c_{j\sigma} + \lambda_v \sum_{i\sigma} \xi_i c^\dagger_{i\sigma} c_{i\sigma}. \tag{41}$$

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

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

$$H_{SO} = i\lambda_{SO} \sum_{\langle\langle ij \rangle\rangle \sigma_1 \sigma_2} \nu_{ij} c^\dagger_{i\sigma_1} s^z_{\sigma_1 \sigma_2} c_{j\sigma_2}. \tag{42}$$

Here $\nu_{ij} = (2/\sqrt{3}) \hat{d}_1 \times \hat{d}_2 = \pm 1$, where i and j are next-nearest-neighbors and $\hat{d}_1$ and $\hat{d}_2$ 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.

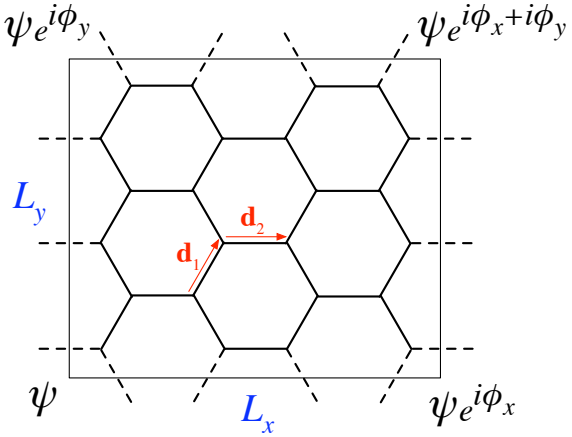

FIG. 1 (Color online) The honeycomb lattice on which the tight-binding Hamiltonian resides. For the two sites depicted, the factor ν_{ij} of equation (42) is $\nu_{ij} = -1$. The phases $\phi_{x,y}$ describe twisted boundary conditions that are used below to give a pumping definition of the $\mathbb{Z}_2$ invariant.

[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.

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 that the square of the time-reversal operator is connected to a 2π rotation, which implies that

$$T^2 = (-1)^{2S}, \tag{43}$$

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

$$\langle T\psi|H'|\psi\rangle = \langle T\psi|H'|T^2\psi\rangle = -\langle T\psi|H'|\psi\rangle = 0, \tag{44}$$

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.

G. 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. Before, in the discussion of the Kosterlitz-Thouless transition, we discussed the behavior of the $U(1)$ nonlinear sigma model, i.e., with the action

$$S_0 = -\frac{K}{2}\int_{\mathbb{R}^2} (\nabla\phi)^2. \tag{45}$$

The direct generalization of this to a more complicated Lie group such as $SU(N)$ is written as

$$S_0 = -\frac{k}{8\pi}\int_{S^2} \mathcal{K}(g^{-1}\partial^\mu g, g^{-1}\partial_\mu g), \tag{46}$$

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

$$S_{WZ} = -\frac{2\pi k}{48\pi^2}\int_{B^3} \epsilon_{\mu\nu\lambda}\mathcal{K}\left(g^{-1}\partial_\mu g, \left[g^{-1}\partial_\nu g, g^{-1}\partial_\lambda g\right]\right) \tag{47}$$

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.

H. 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:

$$\Theta H(k)\Theta^{-1} = H(-k). \tag{48}$$

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:

$$D = \frac{1}{2\pi}\left[\oint_{\partial(EBZ)} dk \cdot \mathcal{A} - \int_{EBZ} d^2k\, \mathcal{F}\right] \mod 2. \tag{49}$$

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, $\mathcal{A}$ 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 (49) 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, (49) 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 (49) 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 (49) 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

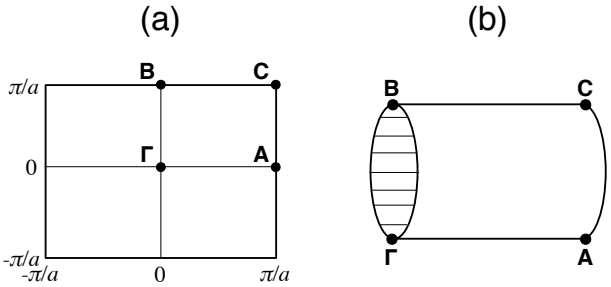

FIG. 2 (a) A two-dimensional Brillouin zone; note that any such Brillouin zone, including that for graphene, can be smoothly deformed to a torus. The labeled points are time-reversal-invariant momenta. (b) The effective Brillouin zone (EBZ). The horizontal lines on the boundary circles ∂(EBZ) connect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

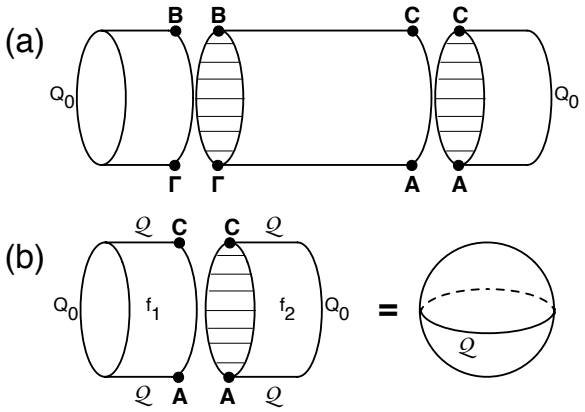

FIG. 3 (a) Contracting the extended Brillouin zone to a sphere. (b) Two contractions can be combined to give a mapping from the sphere, but this sphere has a special property: points in the northern hemisphere are conjugate under $\mathcal{T}$ to those in the southern, in such a way that overall every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

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.

I. 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. [4] Here, the response will again be a pumped charge, but the cyclic process that pumps the chage is more subtle.

Instead of inserting a 2π flux through a circle of the toroidal system, we insert a π flux, adiabatically; this is

[4] 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.

consistent with the part of D in (49) 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 (49), 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.

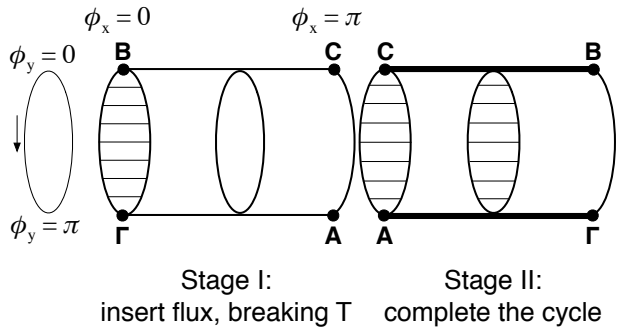

FIG. 4 Graphical representation of charge pumping cycle for Chern parities. The first stage takes place as the flux ϕ_x increases adiabatically from 0 to π. In the second stage the Hamiltonian at $(\phi_x = \pi, \phi_y)$ is adiabatically transported through the space of Hamiltonians to return to the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

J. Experimental status

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 (König et al., Science 2007). 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 set of notes, 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.

II. TOPOLOGICAL PHASES I: THOULESS PHASES ARISING FROM BERRY PHASES, CONTINUED

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.

A. 3D band structure invariants and topological insulators

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 (from Avron, Seiler, and Simon, 1983) 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:

$$x_0 x_{\pm 1} = y_0 y_{\pm 1} = z_0 z_{\pm 1}. \tag{50}$$

(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 $-k$.

B. 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

$$\Delta\mathcal{L}_{EM} = \frac{\theta e^2}{2\pi h}\mathbf{E}\cdot\mathbf{B} = \frac{\theta e^2}{16\pi h}\epsilon^{\alpha\beta\gamma\delta}F_{\alpha\beta}F_{\gamma\delta}. \tag{51}$$

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 PRL 1987) is that, when the axion field $\theta(x,t)$ 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 $\theta(x)$ appear.

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

$$\sigma_{xy} = \frac{e^2(\Delta\theta)}{2\pi h}. \tag{52}$$

(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, see Wilczek, 1987). 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:

$$\theta = \frac{1}{2\pi}\int_{\mathrm{BZ}} d^3k\ \epsilon_{ijk}\,\mathrm{Tr}[\mathcal{A}_i\partial_j\mathcal{A}_k - i\frac{2}{3}\mathcal{A}_i\mathcal{A}_j\mathcal{A}_k], \tag{53}$$

which can be done by imitating our previous derivation of the polarization formula; for details see either Qi, Hughes, Zhang (2008) or Essin, Moore, Vanderbilt (2008). 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,

$$\mathcal{F}_{ij} = \partial_i\mathcal{A}_j - \partial_j\mathcal{A}_i - i[\mathcal{A}_i,\mathcal{A}_j] \tag{54}$$

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

$$dK = \mathrm{Tr}[\mathcal{F}\wedge\mathcal{F}], \tag{55}$$

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

	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 I Comparison of Berry-phase theories of polarization and magnetoelectric polarizability.

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 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

$$\pi_3(SU(N)) = \mathbb{Z}, \quad N \geq 2. \tag{56}$$

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

$$P_i = e\int_{BZ}\frac{d^3k}{(2\pi)^3}\mathrm{Tr}\,\mathcal{A}_i\,. \tag{57}$$

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 (57) in this direction is one charge per transverse unit cell area, i.e.,

e/Ω. Then the ambiguity in dP/dB is

$$\Delta\frac{P_x}{B_x} = \frac{e/\Omega}{h/(e\Omega)} = \frac{e^2}{h} = 2\pi\frac{e^2}{2\pi h}. \tag{58}$$

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.

C. 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. Here we will concentrate on 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.

Remarkably, the AHE originates from a term in the semiclassical equations of motion that is neglected in almost all textbooks. This term was first obtained by Karplus and Luttinger, but as this took place well before the modern idea of Berry phases, their results were not universally accepted. We will present a modern derivation of the Karplus-Luttinger term using the same idea as in our polarization calculation: trying to "gauge away" the Berry phase leads to a gauge-invariant physical effect. We will derive in this process the zero-B-field limit of the standard semiclassical equations of motion in, e.g., Ashcroft and Mermin,

$$\begin{aligned} \hbar\dot{\mathbf{k}} &= e\mathbf{E} + e\mathbf{v}\times\mathbf{B} \\ \hbar\mathbf{v} &= \nabla_k\epsilon_n(\mathbf{k}) + \ldots. \end{aligned} \tag{59}$$

where ... indicate the Karplus-Luttinger term that we seek.

In class, we derived this term for an applied electric field in the form

$$\hbar\mathbf{v} = \nabla_k\epsilon_n\mathbf{k} - e\mathbf{E}\times(\nabla_k\times\mathcal{A}^{(n)}), \tag{60}$$

where $\mathcal{A}^{(n)}$ is the Berry vector potential of band n. The physical interpretation is fairly straightforward once we recall that our polarization calculation already showed that $\mathcal{A}$ can be connected to the spatial distribution of the electron. As an electron wavepacket moves in k-space under the influence of an applied field, there are two contributions to its spatial velocity. The Karplus-Luttinger contribution describes how a change in $\mathbf{k}$ induces a change in the real-space location because the Bloch states are changing; the first term describes how a fixed wavepacket of Bloch states still describes a moving particle.

(Our treatment followed closely chapter 4 of a tutorial by Ong and Lee that is on the course web page, so I will not reiterate it here. One note on their presentation: I believe that deriving their equation (17) requires assuming that the potential is weakly varying, because otherwise higher-order terms are not guaranteed to vanish because of some nonzero commutators.)

FQHE background: in class we gave some standard background on the fractional quantum Hall effect. 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 centered on the Laughlin wavefunction for two-dimensional electrons ($z_j = x_j + iy_j$ describes the jth electron, $j = 1,\ldots,N$)

$$\Psi_m = \left(\prod_{i<j}(z_i - z_j)^m\right) e^{-\sum_i |z_i|^2/4\ell^2}. \tag{61}$$

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,

$$\psi_m = z^m e^{-|z|^2/4\ell^2}. \tag{62}$$

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.

We explained the origin of this wavefunction 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

$$\Psi_{\text{quasihole}} = \left(\prod_i (z_i - z_0) \right) \Psi_m. \tag{63}$$

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.

III. WEN-TYPE TOPOLOGICAL PHASES: THE FRACTIONAL QUANTUM HALL EFFECT

A. 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

$$e^{i\theta} = e^{iq\Phi/(\hbar c)}. \tag{64}$$

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" (cf. Zhang, Hansson, and Kivelson, PRL 1988). [5]

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

$$\mathcal{L} = 2\gamma \epsilon^{\mu\nu\lambda} a_\mu \partial_\nu a_\lambda + a_\mu j^\mu \tag{65}$$

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

$$\epsilon^{\mu\nu\lambda} a_\mu \partial_\nu a_\lambda \to \epsilon^{\mu\nu\lambda} a_\mu \partial_\nu a_\lambda + \epsilon^{\mu\nu\lambda} \partial_\mu \chi \partial_\nu a_\lambda, \tag{66}$$

---

[5] 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.

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

$$\delta S = 2\gamma \int d^2x dt \epsilon^{\mu\nu\lambda} \partial_\mu (\chi \partial_\nu a_\lambda), \tag{67}$$

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

$$4\gamma \epsilon^{\mu\nu\lambda} \partial_n u a_\lambda = -j^\mu. \tag{68}$$

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,

$$\int d^2x (\partial_1 a_2 - \partial_2 a_1) = -\frac{1}{4\gamma} \int d^2x \, j^0. \tag{69}$$

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

$$\theta = \frac{1}{8\gamma}, \tag{70}$$

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 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". [6]

B. 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 (65) 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:

$$J^\mu = \frac{1}{2\pi} \epsilon^{\mu\nu\lambda} \partial_\nu a_\lambda. \tag{71}$$

(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

[6] 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.

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:

$$\mathcal{L}_{CS} = \frac{k}{4\pi}\epsilon^{\mu\nu\lambda} a_\mu \partial_\nu a_\lambda. \tag{72}$$

The point of the new normalization $k = 8\pi\gamma$ compared to (65) 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 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

$$\mathcal{L} = \phi \mathcal{Q} \phi + \phi J, \tag{73}$$

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

$$2\mathcal{Q}\phi = -J \tag{74}$$

which is solved by

$$\phi = \frac{-1}{2\mathcal{Q}} J. \tag{75}$$

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

$$\mathcal{L} = \frac{1}{4} J \frac{1}{\mathcal{Q}} J - J \frac{1}{2\mathcal{Q}} J = -J \frac{1}{4\mathcal{Q}} J. \tag{76}$$

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

$$(\epsilon^{\mu\nu\lambda}\partial_\nu)(\epsilon^{\lambda\alpha\beta}\partial_\alpha a_\beta) = \epsilon^{\mu\nu\lambda}\epsilon^{\lambda\alpha\beta}(\partial_\nu \partial_\alpha a_\beta). \tag{77}$$

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

$$-\partial_\nu^2 a_\mu. \tag{78}$$

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 (65) with the gauge field integrated out is just

$$\mathcal{L} = \frac{1}{8\gamma} j_\mu \left(\frac{\epsilon^{\mu\nu\lambda}\partial_\nu}{\partial^2}\right) j_\lambda. \tag{79}$$

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:

$$\mathcal{L} = \frac{k}{4\pi}\epsilon^{\mu\nu\lambda} a_\mu \partial_\nu a_\lambda - \frac{1}{2\pi}\epsilon^{\mu\nu\lambda} A_\mu \partial_\nu a_\lambda = \frac{k}{4\pi}\epsilon^{\mu\nu\lambda} a_\mu \partial_\nu a_\lambda - \frac{1}{2\pi}\epsilon^{\mu\nu\lambda} a_\mu \partial_\nu A_\lambda, \tag{80}$$

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 (71.

Now we can integrate out a_μ using equation (79) above, recalling $\gamma = k/(8\pi)$, and obtain

$$\mathcal{L}_{\text{eff}} = \frac{\pi}{k} J_\mu \epsilon^{\mu\nu\lambda} \partial_\nu \frac{1}{\partial^2} J_\lambda = \frac{1}{4\pi k} \epsilon^{\mu\alpha\beta} \partial_\alpha A_\beta \epsilon^{\mu\nu\lambda} \partial_\nu \frac{1}{\partial^2} \epsilon^{\lambda\gamma\delta} \partial_\gamma A_\delta. \tag{81}$$

where in the second step we have used the rewritten Lagrangian in (80) 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,

$$\mathcal{L}_{\text{eff}} = \frac{1}{4\pi k} \epsilon^{\mu\nu\lambda} A_\mu \partial_\nu A_\lambda. \tag{82}$$

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

$$J^\mu = -\frac{\delta \mathcal{L}_{\text{eff}}}{\delta A_\mu} = \frac{1}{2\pi k} \epsilon^{\mu\nu\lambda} \partial_\nu A_\lambda. \tag{83}$$

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

$$\sigma_{xy} = \frac{e^2}{(2\pi)k\hbar} = \frac{1}{k}\frac{e^2}{h}. \tag{84}$$

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):

$$J^0 = \delta n = \frac{1}{2\pi k} \delta B. \tag{85}$$

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.

C. 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

$$\mathcal{L}_{CS} = \frac{k}{4\pi} \epsilon^{\mu\nu\lambda} a_\mu \partial_\nu a_\lambda \tag{86}$$

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

$$\mathcal{L} = -\frac{k}{4\pi} \epsilon^{ij} a_i \dot{a}_j, \tag{87}$$

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

$$\epsilon_{ij}\partial_i a_j = 0. \tag{88}$$

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 (87) 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:

$$\theta = \frac{2\pi n_1 a_1}{L_1} + \frac{2\pi n_2 a_2}{L_2}. \tag{89}$$

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

$$S = \int d^2x\, dt\, \frac{k}{4\pi}\epsilon^{\mu\nu\lambda}a_\mu\partial_\nu a_\lambda = -\frac{kL_1L_2}{4\pi}\int dt\, \frac{(2\pi)^2}{L_1L_2}\epsilon^{ij}q_i\dot{q}_j. \tag{90}$$

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.)

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,

$$S_{\rm edge} = \frac{k}{4\pi}\int dt\, dx\, (\partial_t + v\partial_x)\phi\partial_x\phi. \tag{91}$$

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 (67) 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. [7] Then $\epsilon^{ij}a_j = 0$ and we can write $a_i = \partial_i\phi$. Substituting this into the bulk Chern-Simons Lagrangian

$$\begin{aligned} S &= -\frac{k}{4\pi}\int \epsilon^{ij}a_i\partial_0 a_j\, d^2x\, dt = -\frac{k}{4\pi}\int (\partial_x\phi\partial_0\partial_y\phi - \partial_y\phi\partial_0\partial_x\phi)\, d^2x\, dt \\ &= -\frac{k}{4\pi}\int (\partial_x(\phi\partial_0\partial_y\phi) - \partial_y(\phi\partial_0\partial_x\phi))\, d^2x\, dt \\ &= -\frac{k}{4\pi}\int (\nabla\times\mathbf{v})_z\, d^2x\, dt = -\frac{k}{4\pi}\int \mathbf{v}\cdot d\mathbf{l}\, dt, \end{aligned} \tag{92}$$

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

$$\mathbf{v} = (\phi\partial_0\partial_x\phi, \phi\partial_0\partial_y\phi). \tag{93}$$

(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,

$$S_{\rm edge} = \frac{k}{4\pi}\int \partial_t\phi\,\partial_x\phi\, dx\, dt. \tag{94}$$

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, 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

$$S_{\rm edge} = \frac{k}{4\pi}\int (\partial_t\phi - v\partial_x\phi)\,\partial_x\phi\, dx\, dt. \tag{95}$$

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

$$\mathcal{H} = \frac{kv}{4\pi}(\partial_x\phi)^2 \tag{96}$$

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.)

---

[7] 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.

D. 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. [8]. 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 factor of k^{-1} from the quasiparticle propagator since k appears as a coefficient in the Lagrangian. The result is

$$G(q,\omega) \propto \frac{(vq+\omega)^{k-1}}{vq-\omega}. \tag{97}$$

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

$$\mathcal{L} = \frac{1}{4\pi} K^{IJ} a^I_\mu \partial_\nu a^J_\lambda \tag{98}$$

This effective theory works for all but a few proposed quantum Hall states; we will discuss these exotic "non-Abelian" quantum Hall states later.

Appendix A: Physical preliminaries

1. Landau theory of broken-symmetry phases

Most phases of matter, including solids, magnets, superconductors, superfluids, and many others, can be understood in terms of "broken symmetry". At high temperature, fluctuations are induced by the requirement to maximize entropy, and these fluctuations tend to destroy order. As temperature is lowered, the energy gain from developing order can overwhelm the entropy gain from disorder. A remarkable fact that only became clear after the solution of the two-dimensional Ising model by Onsager in 1948 is that this energy-entropy competition can lead to a sharp phase transition, described mathematically by a singularity in some derivative of the free energy that emerges in the thermodynamic limit (the limit of an infinite number of degrees of freedom).

---

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

For our purposes, we need a way to describe such a breaking of symmetry mathematically. Rather than try to describe every microscopic degree of freedom in a complicated interacting system, we will eventually follow Landau and introduce a classical field theory in terms of some emergent or coarse-grained field that describes the type of order we wish to study. To start, let us first consider an Ising model on a hypercubic lattice. Our microscopic description is in terms of a discrete spin variable $s_i = \pm 1$ at each vertex, with the energy function

$$E = -J \sum_{\langle ij \rangle} s_i s_j. \tag{A1}$$

Here ij are nearest-neighbor sites and J is some interaction strength with units of energy. At temperature $T = \infty$, the system is equally likely to be in any microstate. At $T = 0$, only two microstates occur: one with all spins up and one with all spins down. The surprise is that, if the lattice of spins is in more than one dimension, there is a nonzero temperature T_c, proportional to J, below which the zero-temperature description is qualitatively, but not quantitatively correct. As an explicit example, one can construct the lattice mean-field description of the Ising model, leading to the equation

$$m = \tanh(zJm/(k_B T) + h) = \tanh(\beta z J + h), \tag{A2}$$

where $m = \langle s_i \rangle$ is the average spin, $\beta = 1/k_B T$, z is the number of nearest neighbors, h is a possible external magnetic field (defined to include a factor β), and k_B is Boltzmann's constant; for a derivation, cf. Chaikin and Lubensky. The behavior of this self-consistent equation changes at $k_B T = zJ$, and the mean-field transition temperature is $T_c = zJ/k_B$ or $\beta_c = (zJ)^{-1}$.

Physicists say that the system "breaks symmetry" below T_c and picks out a particular sign of the average spin m, where the angle brackets denote thermal averaging. Mathematicians have more satisfactory definitions (because m strictly speaking is always zero): one can look at either whether there is a nonzero correlation function $\langle s_i s_j \rangle$ as i and j become infinitely far apart, or look for a singularity in some derivative of the free energy (in a first derivative for a first-order transition, in some higher derivative for a second-order transition). Even if m is always zero in terms of the Boltzmann sum, physical systems do actually break symmetry, chiefly for dynamical reasons: for example, a bar magnet of iron will in principle explore the whole phase space and flip its north and south poles, but the time it takes to do so may be larger than the age of the universe. Hence we will mostly be content to discuss broken symmetry as real, e.g., $m \neq 0$ in the Ising model, even if that is somewhat sloppy mathematically.

A powerful way to understand the broken symmetry is in terms of two symmetry groups: G, the high-temperature symmetry group, and H, the residual symmetry group that survives in the low temperature phase. We define the "order parameter manifold" as the quotient $M = G/H$, where "dividing by H" (taking cosets of H in G) means that we identify two elements of G that differ by an element of H; note that this is not in general a group. The notion of an "order parameter" is basic in Landau theory: it is the field we use to model all the complicated microscopic states in terms of one, or a few, macroscopic variables. The idea of the order parameter manifold is that, for many interesting phenomena, we do not care especially about the magnitude of the order parameter itself. We care instead about the set of distinguishable low-temperature states at an arbitrary temperature in the ordered phase, which is exactly M.

One reason the set M is important will become clear when we discuss topological defects in the next chapter. Topological defects in an ordered phase can be classified using mappings from spheres in real space to the order parameter manifold M, i.e., the homotopy groups $\pi_n(M)$. We will explain this result and see a number of examples; two other reasons the manifold M is important in practice is that, since moving from one point in M to another is naïvely a "soft" or massless fluctuation, while changing the magnitude of the order parameter is a "hard" or massive fluctuation, using a field-theory description that involves only the degrees of freedom in M, known for historical reasons as a "nonlinear σ-model", is frequently useful.

Landau theory, which we will continue to discuss in the following section, is not restricted to second-order transitions like those we have described above. Actually one of the most powerful predictions of Landau theory, questioned theoretically in recent years, is that second-order transitions require a change of symmetry: more specifically, the symmetry group H must be a subgroup of G or vice versa. Otherwise, Landau theory predicts a first-order phase transition (first-order transitions are always allowed by symmetry, and can be understood simply as level crossings in the free energy of two different phases).

2. Quantum and statistical path integrals

There are two field theories we will deal with in our treatment of broken-symmetry phases. The first is Landau-Ginzburg theory, which can be understood as a power series and gradient expansion of the energy density in terms of

an order parameter field. For the example of the Ising model, in zero magnetic field we can expand the energy density in even powers of $m(\mathbf{r})$:

$$Z_{\mathrm{LG}} = \int (Dm)\, e^{-\beta \int d^d\mathbf{r}\, a_0 + a_1 m^2 + a_2 m^4 + \ldots + b_0 (\nabla m)^2 + \cdots}. \tag{A3}$$

Here the measure of the integral can be defined more precisely in Fourier space, where omitting high-wavevector components is typically necessary for a sensible theory (in condensed matter, this makes physical sense as a "short-distance cutoff" below which the field m is not meaningful). This is a vastly simplified version of the original problem in at least two ways: we are assuming that the integration over the coarse-grained order parameter field $m(\mathbf{r})$, and we are not doing any microscopic calculation of the coefficients that appear in the expansion.

A remarkable fact is that the above Landau-Ginzburg theory can be not just qualitatively correct but actually exact for some properties, such as "critical exponents" near second-order phase transitions, even without a microscopic calculation of the coefficients. Such properties are referred to as "universal": universal properties depend on symmetry and dimensionality but little else. For example, the liquid-gas critical point in the phase diagram of water has the same critical exponents as the Ising phase transition in three dimensions. As a mathematical example of where universality comes from, the terms beyond a_2 in the above energy turn out not to impact critical exponents and selected other properties, as long as the lower-order coefficients have appropriate signs. We will study one or two examples of critical points later, concentrating on examples where topological considerations are important.

The Landau-Ginzburg theory could have been motivated as a high-temperature expansion: we are justified in concentrating on low powers of m because at high enough temperature the order parameter should not be too large. An alternate field theory, the nonlinear σ-model, can be viewed as an expansion starting from zero temperature. We will concentrate on systems in which the zero-temperature phase breaks a continuous symmetry, so that M is a continuous manifold; this includes, for example, Heisenberg and XY magnets, in which $M = S^2$ and $M = S^1$ respectively, but not Ising magnets, where M is a set with two elements. For an XY magnet [9], we can label a ground state simply by an angle θ between 0 and 2π (the order parameter has a magnitude Δ as well, but all ground states have the same magnitude of the order parameter by symmetry).

When the temperature is slightly increased, fluctuations of the order parameter will take place. The nonlinear σ-model is a theory that ignores fluctuations of the order parameter *magnitude* but captures fluctuations in its *direction*, which are lower in energy or "softer". More precisely, the nonlinear σ-model into a symmetric space $M = G/H$ is defined as a path integral over an M-valued field. For the XY case above, this can be written simply in terms of a spatially varying angle $\theta(\mathbf{r})$:

$$Z_{\mathrm{NL}\sigma\mathrm{M}} = \int (D\theta)\, e^{-\int d^d\mathbf{r}\, g(\nabla\theta)^2}, \tag{A4}$$

where we have incorporated β into the definition of the coupling g. We will return to this model once we have said a bit more about topological defects; it turns out that for our XY example in two spatial dimensions, the physics depends crucially on "vortices", and in fact shows a phase transition that would not be present if, hypothetically, the field θ were not periodic and Eq. A4 became just the Gaussian model.

We have written both of the above field theories in a classical or Euclidean representation, where A natural question is how the partition function integral in such a theory is related to the quantum path integral that may be familiar from an advanced course in quantum mechanics. The easiest example of the analytic continuation to imaginary time that connects the two types of path integrals is for the harmonic oscillator. Its partition function at a finite temperature T is

$$Z_{\mathrm{harmonic}} \approx \int dx(\tau)\, e^{\int_0^\beta d\tau\, \dot{x}^2(\tau)/2m + kx^2/2}, \tag{A5}$$

where there are periodic boundary conditions on $x(\tau)$: $x(\beta) = x(0)$. A worthwhile calculation (hint: simplify the integral by considering Fourier components of $x(\tau)$) leads to the result

$$Z_{\mathrm{harmonic}} = \frac{1}{2\sinh(\beta\hbar\omega/2)} = \sum_{n=0}^{\infty} e^{-\beta\hbar\omega(n+1/2)}, \tag{A6}$$

---

[9] We choose the example of $M = U(1) \cong SO(2)$ here for a reason. It turns out that, for the nonlinear σ-model to include gapless excitations, the form of the theory becomes more complicated. For Lie groups more complicated than $U(1)$, an additional term of topological origin is required, leading to the Wess-Zumino-Novikov Witten model that we discuss in Section .

where the last expression is what we would calculate from the spectrum. Now analytically continuing this calculation from imaginary time τ gives a trace of the

$$Z_{\text{harmonic}} = \text{Tr } e^{-\beta H} \to \text{Tr } e^{-itH/\hbar} = \int dx_0 \, U(x_0, t; x_0, 0), \tag{A7}$$

where in the last step we have used the position basis to put the result in terms of matrix elements of the unitary time evolution operator U. Now the divergence of Z at real times $t = 2\pi n/\omega$, for integer n, can be simply interpreted: at these times all the energy eigenstates that appear in an arbitrary initial condition appear with exactly the same phases, so the state is (aside from an overall phase factor) exactly the initial state, the time evolution operator is the identity, and the trace diverges.