Draft Lecture II notes for Les Houches 2014

Joel E. Moore, UC Berkeley and LBNL

(Dated: August 7, 2014)

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

(For the standard treatment based on Landau levels, we refer the reader to the books by Prange and Girvin, or Das Sarma and Pinczuk.)

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

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{\mathbf{k}}(\mathbf{r}), \tag{1}$$

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

We give a quick proof of Bloch's theorem in one spatial dimension, then consider the Berry phase of the resulting wavefunctions. A standard fact from quantum mechanics tells us that, given two Hermitian operators that commute, we can find a basis of simultaneous wavefunctions. In the problem at hand, we have a non-Hermitian operator (lattice translations by the lattice spacing a: $(T\psi)(x) = \psi(x+a)$) that commutes with the Hamiltonian. It turns out that only one of the two operators needs to be Hermitian for simultaneous eigenstates to exist, and therefore we can find wavefunctions that are energy eigenstates and satisfy

$$(T\psi)(x) = \lambda\psi(x). \tag{2}$$

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

$$\psi_k(x+a) = e^{ika}\psi_k(x). \tag{3}$$

The last step is to define $u_k(x) = \psi_k(x)e^{-ikx}$; then (3) shows that u_k is periodic with period a, and $\psi_k(x) = e^{ikx}u_k(x)$. [1]

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

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"' $k \in [-\pi/a, \pi/a)$:

$$\gamma = \oint_{-\pi/a}^{\pi/a} \langle u_k | i\partial_k | u_k \rangle dk. \tag{4}$$

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

Another clue to what the phase γ might mean physically is provided by asking if it is gauge-invariant. Before, gauge-invariance resulted from assuming that the wavefunction could be continuously defined on the interior of the closed path. Here we have a closed path on a noncontractible manifold; the path in the integral winds around the Brillouin zone, which has the topology of the circle. What happens to the Berry phase if we introduce a phase change $\phi(k)$ in the wavefunctions, $|u_k\rangle \rightarrow e^{-i\phi(k)}|u_k\rangle$, with $\phi(\pi/a) = \phi(-\pi/a) + 2\pi n, n \in \mathbb{Z}$? Under this transformation, the integral shifts as

$$\gamma \rightarrow \gamma + \oint_{-\pi/a}^{\pi/a} (\partial_k \phi)\, dk = \gamma + 2\pi n. \tag{5}$$

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

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

$$n_j = \frac{i}{2\pi} \int dk_x\, dk_y\, \left(\langle \partial_{k_x} u_j | \partial_{k_y} u_j \rangle - \langle \partial_{k_y} u_j | \partial_{k_x} u_j \rangle \right) \tag{6}$$

to the total Hall conductance:

$$\sigma_{xy} = \frac{e^2}{h} \sum_j n_j. \tag{7}$$

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

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:

$$\Delta P = \int_0^1 d\lambda \frac{dP}{d\lambda} = \int_{t_0}^{t_1} dt\, \frac{dP}{d\lambda}\frac{d\lambda}{dt} = \int_{t_0}^{t_1} j(t)\, dt. \tag{8}$$

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

$$H(q,\lambda) = \frac{1}{2m}(p+\hbar q)^2 + V^{(\lambda)}(r). \tag{9}$$

The current operator is

$$j(q) = ev(q) = \frac{ie}{\hbar}[H(q,\lambda), r] = \frac{e}{m}(p+\hbar q) = \frac{e}{\hbar}\partial_q H(q,\lambda). \tag{10}$$

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

$$|\psi(t)\rangle = \exp\left(-(i/\hbar)\int^t E_0(t')\,dt'\right)\left[|\psi_0(t)\rangle + i\hbar\sum_{j\neq 0}|\psi_j(t)\rangle(E_j - E_0)^{-1}\langle\psi_j(t)|\dot{\psi}_0(t)\rangle\right]. \tag{11}$$

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

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

$$\frac{dP}{d\lambda} = 2\hbar e\,\mathrm{Im}\sum_{v,c}\int\frac{dq}{2\pi}\frac{\langle u_v(q)|v(q)|u_c(q)\rangle\langle u_c(q)|\partial_\lambda u_v(q)\rangle}{E_c(q)-E_v(q)}. \tag{12}$$

For example, we wrote

$$\langle\psi_j(t)|\dot{\psi}_0(t)\rangle = \sum_{v,c}\langle u_c|\partial_\lambda u_v\rangle\frac{d\lambda}{dt}. \tag{13}$$

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

$$|\partial_q u_j(q)\rangle = \sum_{j\neq j'}|u_{j'}(q)\rangle\frac{\langle u_{j'}(q)|\partial_q H(q,\lambda)|u_j(q)\rangle}{E_j(q)-E_{j'}(q)}. \tag{14}$$

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

$$\frac{dP}{d\lambda} = 2\hbar e\,\mathrm{Im}\sum_{c}\int\frac{dq}{2\pi}\frac{\langle u_v(q)|v(q)|u_c(q)\rangle\langle u_c(q)|\partial_\lambda u_v(q)\rangle}{E_c(q)-E_v(q)} = 2e\,\mathrm{Im}\int\frac{dq}{2\pi}\langle\partial_q u_v(q)|\partial_\lambda u_v(q)\rangle. \tag{15}$$

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

$$\Delta P = 2e\,\mathrm{Im}\int_0^1 d\lambda\int\frac{dq}{2\pi}\langle\partial_q u_v(q)|\partial_\lambda u_v(q)\rangle. \tag{16}$$

The last expression is in two dimensions and involves the same type of integrand (a Berry flux) as in the 2D TKNN formula (6). However, in the polarization case there does not need to be any periodicity in the parameter λ. If this parameter is periodic, so that $\lambda = 0$ and $\lambda = 1$ describe the same system, then the total current run in a closed cycle that returns to the original Hamiltonian must be an integer number of charges, consistent with quantization of the TKNN integer in the IQHE.

If we define polarization via the Berry connection,

$$P = ie \int \frac{dq}{2\pi} \langle u_v(q) | \partial_q u_v(q) \rangle, \tag{17}$$

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

D. Interactions and disorder: the flux trick

One might worry whether the TKNN integer defined in equation (6) is specific to noninteracting electrons in perfect crystals. An elegant way to generalize the definition physically, while keeping the same mathematical structure, was developed by Niu, Thouless, and Wu. This definition also makes somewhat clearer, together with our polarization calculation above, why this invariant should describe σ_{xy}. First, note that from the formula for the Bloch Hamiltonian in the polarization calculation above, we can reinterpret the crystal momentum q as a parameter describing a flux threaded through a unit cell of size a: the boundary conditions are periodic up to a phase $e^{iqa} = e^{ie\Phi/\hbar c}$. We will start by reinterpreting the noninteracting case in terms of such fluxes, then move to the interacting case.

The setup is loosely similar to the Laughlin argument for quantization in the IQHE. Consider adiabatically pumping a flux Φ_x though one circle of a toroidal system, in the direction associated with the periodicity $x \to x + L_x, y \to y$. The change in this flux in time generates an electric field pointing in the $\hat{\mathbf{x}}$ direction. Treating this flux as a parameter of the crystal Hamiltonian, we compute the resulting change in $\hat{\mathbf{y}}$ polarization, which is related to the y current density:

$$\frac{dP_y}{dt} = j_y = \frac{dP_y}{d\Phi_x}\frac{d\Phi_x}{dt} = \frac{dP_y}{d\Phi_x}(cE_xL_x). \tag{18}$$

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

$$P_y(\Phi_x) = \frac{ie}{2\pi} \int d\Phi_y \, \langle u | \partial_{\Phi_y} u \rangle \tag{19}$$

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

$$\langle j_y \rangle = \langle \frac{dP_y}{d\Phi_x} \rangle (cE_xL_x) \to \frac{\Delta P_y}{\Delta \Phi_x}(cE_xL_x), \tag{20}$$

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

$$\langle j_y \rangle = \frac{e}{hc}\frac{ne}{L_x}(cE_xL_x) = \frac{ne^2}{h}E_x. \tag{21}$$

The integer n follows from noting that computing $dP_y/d\Phi_x$ and then integrating $d\Phi_x$ gives just the expression for the TKNN integer (6), 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

$$n_j = \frac{i}{2\pi} \int \mathrm{Tr}\,[dP_j \wedge P_j\, dP_j]\,, \tag{22}$$

where $\wedge$ is the wedge product and $dP_j = \partial_{k_x} P_j\, dk_x + \partial_{k_y} P_j\, dk_y$ is a differential form where the coefficients are operators. (Note that the wedge product in the above formula acts only on dk_x and dk_y.) It is a straightforward exercise to verify that this reproduces the TKNN definition (6).

Then the generalization to degenerate bands, for example, is naturally studied by using the gauge- and basis-invariant projection operator $P_{ij} = |u_i\rangle\langle u_i| + |u_j\rangle\langle u_j|$ onto the subspace spanned by $|u_i\rangle$ and $|u_j\rangle$: the index of this operator gives the total Chern number of bands i and j. In general, when two bands come together, only their total Chern number is defined. The total Chern number of all bands in a finite-dimensional band structure (i.e., a finite number of bands) is argued to be zero below. Often one is interested in the total Chern number of all occupied bands because this describes the integer quantum Hall effect through the TKNN formula; because of this zero sum rule, the total Chern number of all *unoccupied* bands must be equal and opposite.

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

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

$$H(k_x, k_y) = a_0(k_x, k_y)\mathbf{1} + a_1(k_x, k_y)\sigma_x + a_2(k_x, k_y)\sigma_y + a_3(k_x, k_y)\sigma_z. \tag{23}$$

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

$$(a_1, a_2, a_3) \to \frac{(a_1, a_2, a_3)}{\sqrt{{a_1}^2 + {a_2}^2 + {a_3}^2}} \tag{24}$$

at each point in k-space.

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

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

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

[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.

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:

$$H(k) = U(k)D(k)U^{-1}(k). \tag{25}$$

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

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

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 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

$$J^i_j = \alpha_s \epsilon_{ijk} E_k, \tag{26}$$

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

As an example of this "two copies of the IQHE" generated by spin-orbit coupling, consider the model of graphene introduced by Kane and Mele.(**?**) This is a tight-binding model for independent electrons on the honeycomb lattice (Fig. 1). The spin-independent part of the Hamiltonian consists of a nearest-neighbor hopping, which alone would

[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.

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:

$$H_0 = t \sum_{\langle ij \rangle \sigma} c^\dagger_{i\sigma} c_{j\sigma} + \lambda_v \sum_{i\sigma} \xi_i c^\dagger_{i\sigma} c_{i\sigma}. \tag{27}$$

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

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

$$H_{SO} = i\lambda_{SO} \sum_{\langle\langle ij \rangle\rangle \sigma_1 \sigma_2} \nu_{ij} c^\dagger_{i\sigma_1} s^z_{\sigma_1\sigma_2} c_{j\sigma_2}. \tag{28}$$

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.

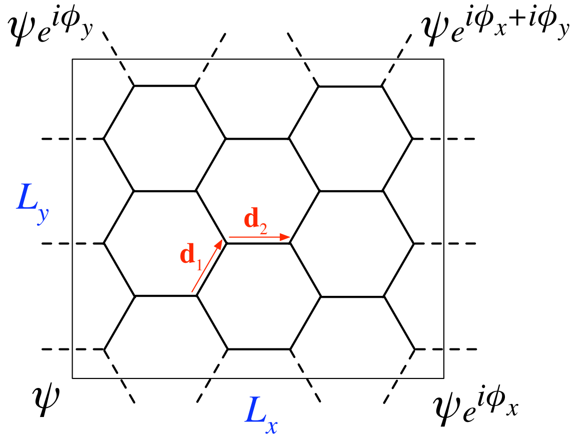

FIG. 1 (Color online) The honeycomb lattice on which the tight-binding Hamiltonian resides. For the two sites depicted, the factor ν_{ij} of equation (28) is $\nu_{ij} = -1$. The phases $\phi_{x,y}$ describe twisted boundary conditions that are used below to give a pumping definition of the $\mathbb{Z}_2$ invariant.

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

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

$$T^2 = (-1)^{2S}, \tag{29}$$

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

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

$$\langle T\psi | H' | \psi \rangle = \langle T\psi | H' | T^2 \psi \rangle = -\langle T\psi | H' | \psi \rangle = 0, \tag{30}$$

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

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

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

A direct Abelian Berry-phase approach for the 2D $\mathbb{Z}_2$ invariant is provided in the Appendix, along with an introduction to Wess-Zumino terms in 1+1-dimensional field theory and a physical interpretation of the invariant in terms of pumping cycles. The common aspect between these two is that in both cases the "physical" manifold (either the 2-sphere in the WZ case, or the 2-torus in the QSHE case) is extended in a certain way, with the proviso that the resulting physics must be independent of the precise nature of the extension. When we go to 3 dimensions in the following lecture, it turns out that there is a very nice 3D non-Abelian Berry-phase expression for the 3D $\mathbb{Z}_2$ invariant; while in practice it is certainly no easier to compute than the original expression based on applying the 2D invariant, it is much more elegant mathematically so we will focus in that. Actually, for practical calculations, a very important simplification for the case of inversion symmetry (in both $d = 2$ and $d = 3$) was made by Fu and Kane: the topological invariant is determined by the product of eigenvalues of the inversion operator at the 2^d time-reversal symmetric points of the Brillouin zone. For further details we refer the reader to their 2007 PRB.

G. Experimental status of 2D insulating systems

This completes our discussion of one- and two-dimensional insulating systems. The two-dimensional topological insulator was observed by a transport measurement in $(Hg, Cd)Te$ quantum wells (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.

Topological invariants in 2D with time-reversal invariance

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

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

$$S_0 = -\frac{K}{2} \int_{\mathbb{R}^2} (\nabla \phi)^2. \tag{31}$$

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

$$S_0 = -\frac{k}{8\pi} \int_{S^2} \mathcal{K}(g^{-1}\partial^\mu g, g^{-1}\partial_\mu g), \tag{32}$$

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

$$S_{WZ} = -\frac{2\pi k}{48\pi^2} \int_{B^3} \epsilon_{\mu\nu\lambda} \mathcal{K}\left(g^{-1}\partial_\mu g, \left[g^{-1}\partial_\nu g, g^{-1}\partial_\lambda g\right]\right) \tag{33}$$

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.

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

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

$$\Theta H(k) \Theta^{-1} = H(-k). \tag{34}$$

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:

$$D = \frac{1}{2\pi}\left[\oint_{\partial(EBZ)} dk \cdot \mathcal{A} - \int_{EBZ} d^2k \, \mathcal{F}\right] \quad \text{mod } 2. \tag{35}$$

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 (35) 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, (35) 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 (35) 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 (35) is invariant of which contraction we choose. The first is rather straightforward and follows from $\pi_1(\mathcal{H}) = \pi_1(\mathcal{Q}) = 0$. The second step is more subtle and gives an understanding of why only a $\mathbb{Z}_2$ invariant or "Chern parity" survives, rather than an integer-valued invariant as the IQHE. We can combine two different contractions of the same boundary into a sphere, and the Chern number of each band pair on this sphere gives the difference between the Chern numbers of the band pair obtained using the two contractions (Fig. 3).

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

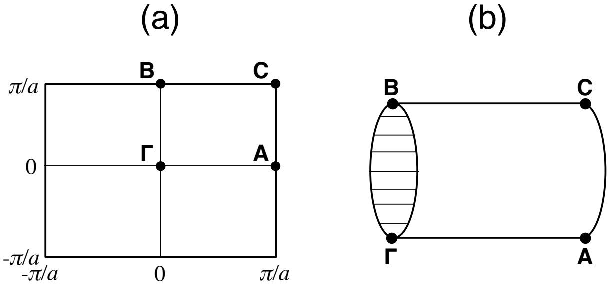

FIG. 2 (a) A two-dimensional Brillouin zone; note that any such Brillouin zone, including that for graphene, can be smoothly deformed to a torus. The labeled points are time-reversal-invariant momenta. (b) The effective Brillouin zone (EBZ). The horizontal lines on the boundary circles ∂(EBZ) connect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

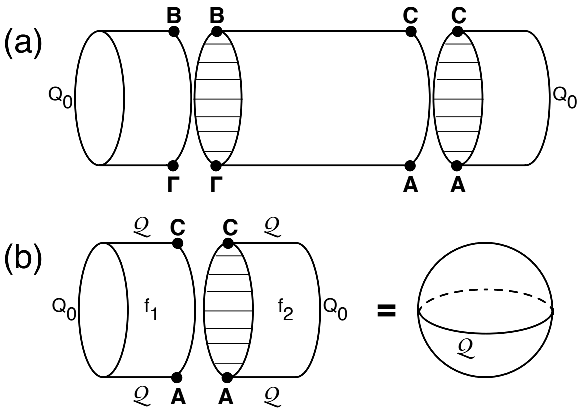

FIG. 3 (a) Contracting the extended Brillouin zone to a sphere. (b) Two contractions can be combined to give a mapping from the sphere, but this sphere has a special property: points in the northern hemisphere are conjugate under $\mathcal{T}$ to those in the southern, in such a way that overall every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

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 ↑, $\nu = -1$ for spin ↓ 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.

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

We expect that, as for the IQHE, it should be possible to reinterpret the $\mathbb{Z}_2$ invariant as an invariant that describes the response of a finite toroidal system to some perturbation. In the IQHE, the response is the amount of charge that is pumped around one circle of the torus as a 2π flux (i.e., a flux hc/e) is pumped adiabatically through the other circle. [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 consistent with the part of D in (35) 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 (35), 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.

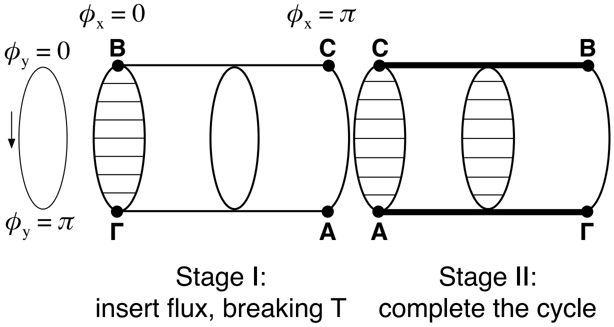

FIG. 4 Graphical representation of charge pumping cycle for Chern parities. The first stage takes place as the flux ϕ_x increases adiabatically from 0 to π. In the second stage the Hamiltonian at $(\phi_x = \pi, \phi_y)$ is adiabatically transported through the space of Hamiltonians to return to the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

[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.