Topological insulators and their invariants

Christian Schmidt
Department of Physics at the University of British Columbia
(Dated: December 19, 2021)

We present an introductory review of the invariants which distinguish topological insulators, and their origins. To this end, we present the theory of Berry phase and curvature, leading to the definition of the first Chern number, the topological invariant of two-dimensional insulators with nondegenerate bands. We follow it up with summary of how to find the number and type of topological invariants in the more general case and quickly introduce the idea of symmetry-protected topological invariants.

I. INTRODUCTION

Topological insulators are distinguished by two extraordinary properties: Robustness against perturbations and the existence of gapless boundary states. Regular perturbations, no matter how strong, cannot change them into topologically trivial systems without closing their band gap. This directly leads to the second property, that a boundary between a topological insulator and a topologically trivial material always exhibits gapless states at the phase boundary [1]. These modes exist independently of the structure of this boundary and regardless of disorder or arbitrarily strong local potentials [2].

A topological material is mathematically characterized by a nonzero topological invariant. The simplest and most instructive case is the first Chern number for a two-dimensional system, which can be constructed through the theory of Berry phase and curvature. We will lay out this construction in detail, and follow it up with a summary of how to find the number and type of topological invariants in the more general case.

II. BERRY PHASE

Consider a system with its Hamiltonian $H(\boldsymbol{k})$ smoothly dependent on some parameter $\boldsymbol{k}$. In our context of topological insulators, this parameter is momentum, although the Berry phase can also be defined with other parameters like external fields [11].

When the parameter is changed slowly over time, the adiabatic theorem [3] implies that eigenstates of $H(\boldsymbol{k}(t))$ stay eigenstates. In fact, this can be seen with the ansatz

$$H(\boldsymbol{k}) = \sum_n E_n(\boldsymbol{k}) \left|n(\boldsymbol{k})\right\rangle \left\langle n(\boldsymbol{k})\right| \tag{1}$$

$$\left|\psi(t)\right\rangle = \sum_n c_n(\boldsymbol{k}(t))\, e^{\int E_n \mathrm{d}t} \left|n(\boldsymbol{k}(t))\right\rangle \tag{2}$$

where E_n and $|n\rangle$ are the eigenvectors of the Hamiltonian at a specific $\boldsymbol{k}$, and c_n are the components of $|\psi\rangle$ in the given basis. Applying the Schrödinger equation for the time evolution of $|\psi\rangle$ yields

$$i\frac{\mathrm{d}}{\mathrm{d}t}\left|\psi(t)\right\rangle = H(\boldsymbol{k}(t))\left|\psi(t)\right\rangle \tag{3}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}c_n = -\left\langle n\right| \frac{\mathrm{d}}{\mathrm{d}t}\left|n\right\rangle c_n + \mathcal{O}\left(\frac{\frac{\mathrm{d}}{\mathrm{d}t}H}{|E_m - E_n|}\right) \tag{4}$$

Adiabaticity means that the change of the Hamiltonian is very slow compared to the smallest gap between eigenvalues (which requires non-degeneracy at all $k(t)$). If this is satisfied, we get the solution

$$c_n(t) = e^{-\int \left\langle n\right| \frac{\mathrm{d}}{\mathrm{d}t}\left|n\right\rangle \mathrm{d}t} c_n(0) \tag{5}$$

Note that $\langle n| \frac{\mathrm{d}}{\mathrm{d}t} |n\rangle$ is imaginary since

$$2\,\mathrm{Re}\left\langle n\right| \frac{\mathrm{d}}{\mathrm{d}t}\left|n\right\rangle = \frac{\mathrm{d}}{\mathrm{d}t}\left\langle n|n\right\rangle = 0 \tag{6}$$

so the exponential term describes a complex phase.

The definition of $|n(\boldsymbol{k})\rangle$ and consequently the values of $c_n(\boldsymbol{k})$ and ϕ^B contain an ambiguity: You can "regauge" the phases as

$$\left|\tilde{n}(\boldsymbol{k})\right\rangle = e^{i\chi(\boldsymbol{k})}\left|n(\boldsymbol{k})\right\rangle \tag{7}$$

$$\tilde{c}_n(\boldsymbol{k}) = e^{-i\chi(\boldsymbol{k})} c_n(\boldsymbol{k}) \tag{8}$$

with no change to H or $|\psi\rangle$. This means that the phases are not physical properties of the Hamiltonian, but can be arbitrarily chosen; this is called a *gauge freedom*. However, a change in gauge cannot always absorb phases in $c_n(t)$: If the path $k(t)$ returns to a previous value, as e.g. $k(T) = k(0)$, a phase change in $|n(\boldsymbol{k}(0)\rangle$ affects both $c_n(0)$ and $c_n(T)$ equally. In this case, a phase change cannot modify the value of

$$\phi^B = \arg\left(\frac{c_n(t)}{c_n(0)}\right) \tag{9}$$

$$= \int_0^t \left\langle n\right| i\frac{\mathrm{d}}{\mathrm{d}t}\left|n\right\rangle \mathrm{d}t \mod 2\pi \tag{10}$$

This is called the *Berry phase* [4] along the curve $\mathcal{C} : t \mapsto \boldsymbol{k}(t)$ and energy level n. A nonzero Berry phase means that if the parameter is slowly changed to describe the closed curve C, the time evolution brings an initial state $|n(\boldsymbol{k})\rangle$ to a final state proportional to the initial

one (as required by the adiabatic theorem), but with its phase changed by ϕ^B In fact, this phase only depends on the curve and not on its parameterization by t:

$$\phi^B = \int_0^t \langle n|\, i\nabla_{\boldsymbol{k}}\, |n\rangle \cdot \frac{\mathrm{d}\boldsymbol{k}}{\mathrm{d}t}\,\mathrm{d}t \mod 2\pi \tag{11}$$

$$= \oint_{\mathcal{C}} \langle n|\, i\nabla_{\boldsymbol{k}}\, |n\rangle \cdot \mathrm{d}\boldsymbol{k} \mod 2\pi \tag{12}$$

Due to this property, the Berry phase is called a *geometric* phase.

III. BERRY CURVATURE

The Berry phase has the form

$$\phi^B = \oint_{\mathcal{C}} \mathcal{A}(\boldsymbol{k}) \cdot \mathrm{d}\boldsymbol{k} \mod 2\pi \tag{13}$$

$$\mathcal{A}(\boldsymbol{k}) = \langle n|\, i\nabla_{\boldsymbol{k}}\, |n\rangle \tag{14}$$

where $\mathcal{A}$ is called the *Berry connection*. This form suggests that we can apply the generalized Stokes' theorem to express ϕ^B as a surface integral:

$$\phi^B = \int_{\mathcal{S}} \sum_{ij} \mathcal{F}_{ij}(\boldsymbol{k})\, \mathrm{d}\boldsymbol{k}_i \wedge \mathrm{d}\boldsymbol{k}_j \mod 2\pi \tag{15}$$

$$\mathcal{F}_{ij}(\boldsymbol{k}) = \partial_i \mathcal{A}_j(\boldsymbol{k}) - \partial_j \mathcal{A}_i(\boldsymbol{k}) \tag{16}$$

where $\mathcal{S}$ is a surface such that $\mathcal{C}$ is its boundary. The tensor $\mathcal{F}$ is called the *Berry curvature* [4]. For the case of 3-dimensional reciprocal space, its definition reduces to

$$\mathcal{F}(\boldsymbol{k}) = \nabla_{\boldsymbol{k}} \times \mathcal{A}(\boldsymbol{k}) \tag{17}$$

The Berry curvature describes the behavior of the Berry phase for very small curves: If $\mathcal{C}$ encloses an infinitesimal area $\mathrm{d}S$, the Berry phase is $\mathrm{d}\phi^B = \mathcal{F} \cdot \mathrm{d}S$

Analyzing the behavior of $\mathcal{A}$ and $\mathcal{F}$ under gauge transformations (changes in phase convention), we get

$$|\tilde{n}(\boldsymbol{k})\rangle = e^{i\chi(\boldsymbol{k})}\, |n(\boldsymbol{k})\rangle \tag{18}$$

$$\tilde{\mathcal{A}}(\boldsymbol{k}) = \mathcal{A}(\boldsymbol{k}) + \nabla_{\boldsymbol{k}}\, \chi(\boldsymbol{k}) \tag{19}$$

$$\tilde{\mathcal{F}}(\boldsymbol{k}) = \mathcal{F}(\boldsymbol{k}) \tag{20}$$

This means that the expression for ϕ^B in terms of $\mathcal{F}$ no longer needs to be taken modulo 2π to be independent of the phase convention.

IV. CHERN NUMBERS

Consider the expression

$$2\pi C_1 = \int_{\mathcal{S}} \sum_{ij} \mathcal{F}_{ij}(\boldsymbol{k})\, \mathrm{d}\boldsymbol{k}_i \wedge \mathrm{d}\boldsymbol{k}_j \tag{21}$$

for a closed surface $\mathcal{S}$ in $\boldsymbol{k}$-space. When taken modulo 2π, it gives the Berry phase (eq. 16) along the boundary curve of $\mathcal{S}$ - but $\mathcal{S}$ has no boundary, so we know that this integral gives a multiple of 2π. In other words, the *first Chern number* C_1 is an integer - but it is not necessarily zero. [5]

For a two-dimensional system, the Brillouin zone itself forms a closed surface, specifically a two-dimensional torus. Setting $\mathcal{S}$ to the full Brillouin zone thus gives a Chern number for the system itself.

Here we have defined the *first* Chern number. The same kind of construction can be applied to $2n$-dimensional systems, by integrating higher polynomials in the curvature $\mathcal{F}$ over the parameter space, to get the nth Chern number [6].

The first Chern number as defined by the surface integral (eq. 21) is a continuous function of $\mathcal{F}$, which in turn depends continuously on the eigenstates $|n\rangle$. However, any continuous function which always gives integer values must be constant - so no regular perturbation to the Hamiltonian can change the Chern number. Due to this property, the Chern number is called a *topological* invariant.

V. TOPOLOGICALLY PROTECTED STATES

The significance of the Chern number comes precisely from the fact that a change in Chern number requires a discontinuous transition between band structures.

This is also true for a boundary between materials with different topological invariants: Any crossover from a material with nonzero Chern number to a topologically trivial state (with $C_1 = 0$) requires breaking one of our earliest assumptions, namely that the bands are nondegenerate everywhere. At some point along the transition, the Chern number must jump from one value to another. This can only happen if the bands become degenerate there.

For a topological insulator (with a Fermi energy between two of the bands in question) this forces gapless edge states to exist [2]. This argument does not depend in any way on the structure of the boundary, so the edge states cannot be suppressed/"localized" by disorder or strong potentials, which is very unusual.

We will not go into the details of the edge states here, since that is not the focus of this project.

VI. DEGENERATE CASE

It is not very often the case that valence and conduction bands in a topological insulator are non-degenerate everywhere, as required for the preceding discussion. However, as long as the system is an insulator and thus has a band gap at the Fermi level, we can separate all bands into occupied and unoccupied states, with no adiabatic cross-over between them. In the remainder of this

section, we will trace the steps needed to calculate the group of topological invariants in this case.

The band structure of a $n \times n$ Hamiltonian with band gap at $E = 0$ can be specified as a smooth map from the Brillouin zone BZ (with periodic boundary) to a set of n nonzero eigenvalues and n orthonormal eigenvectors. From a topological viewpoint, the energy levels can be continuously deformed to any value, as long as they do not cross the band gap, so they only divide the states into two distinct sets of k occupied eigenvectors with an associated eigenvalue $\epsilon < 0$ and $(n-k)$ unoccupied ones with $\epsilon > 0$. Similarly, a continuous change can replace the eigenvectors by (orthonormality-preserving) linear combinations of vectors within the same set.

An ordered collection of n eigenvalues can be represented as the columns of an unitary $n \times n$-matrix, and the allowed transformations on the occupied and unoccupied subsets are unitary $k \times k$ or $(n-k) \times (n-k)$ group actions. The resulting quotient space is the *Grassmannian* $\mathbf{Gr}(k, n) = U(n)/(U(k) \times U(n-k))$. A way to think about elements of the Grassmannian is the k-dimensional subspace of $\mathbb{C}^n$ spanned by the occupied states.

This shows that the space of all gapped Hamiltonians with d-dimensional Brillouin zone BZ and k occupied, $(n-k)$ unoccupied bands is homotopy equivalent[12] to the space of maps $\mathrm{BZ} \to \mathbf{Gr}(k, n)$. If k and $(n-k)$ are sufficiently high, the set of such maps, up to homotopy, can be calculated as the homotopy groups

$$\pi_d(\mathbf{Gr}(k,n)) = \begin{cases} \mathbb{Z}, & d \text{ even} \\ \{\mathrm{e}\}, & d \text{ odd} \end{cases} \tag{22}$$

Topological invariants can discriminate only between Hamiltonians which are not homotopic, so the set of possible values for a topological invariant is equivalent to the set of band structures up to homotopy.

The previous calculation thus shows that in an even number of dimensions, there is one integer-valued topological invariant which classifies band-gapped Hamiltonians, and in odd dimensions there is no such invariant. The invariant in even dimensions is the Chern number discussed in section IV. [2]

VII. SYMMETRY-PROTECTED TOPOLOGY

Many physical systems have discrete symmetries such as time-reversal. Such a symmetry reduces the space of possible Hamiltonians and band structures, which can completely change its topology. It can both disallow topologically nontrivial systems that were previously possible, but also remove systems which could normally interpolate between remaining Hamiltonians to create new topological invariants. If a material has a nonzero value for such a new invariant, it is a *symmetry-protected topological material*.

Symmetry-protected topological insulators show gapless boundary states just like in the symmetry-less case. However, the symmetry-protected topological phase differs in that it is robust against perturbations only if the perturbation respects the original symmetry of the system.

There are three discrete symmetries relevant for this discussion: time-reversal, particle-hole and chiral symmetry, the latter of which is the composition of time reversal and particle-hole symmetry operations.

Through the same procedure as presented in the previous section, one can find the groups of topological invariants for all the different restrictions on the band structure. They were first tabulated by Zirnbauer and Altland [9] [10], a review for the case of up to 3 dimensions can be found in [2].

VIII. CONCLUSION

Starting from the adiabatic evolution of a state, we have defined the *Berry phase*, which describes the phase change along a closed loop in momentum space. Writing this phase as an integral over the enclosed surface and integrating over the whole Brillouin zone, we arrived at the *first Chern number*, which can only take integer values. As a topological invariant, it is unaffected by any regular perturbations and it can only change values at a boundary by producing gapless states.

We have derived the structure of possible topological invariants, showing that in the absence of symmetry, topological insulators only exist in even dimensions and there is one integer as an invariant to discern between topologically distinct states. Taking into account discrete symmetries, we ended with a discussion of symmetry-protected topological insulators, which only retain their topological properties as long as the symmetry is preserved.

[1] Kane, C. L.; Mele, E. J., "Z2 Topological Order and the Quantum Spin Hall Effect", Physical Review Letters. 95 (14): 146802 (2005) https://doi.org/10.1103/PhysRevLett.95.146802

[2] Andreas P. Schnyder, Shinsei Ryu, Akira Furusaki, and Andreas W. W. Ludwig , "Classification of Topological Insulators and Superconductors", AIP Conference Proceedings 1134, 10-21 (2009) https://doi.org/10.1063/1.3149481

[3] Born, M. and Fock, V., "Beweis des Adiabatensatzes", Z. Physik, 51, 165-180 (1928). https://doi.org/10.1007/BF01343193

[4] Berry, M. V. “Quantal Phase Factors Accompanying Adiabatic Changes.” Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences 392, no. 1802 (1984): 45–57. http://www.jstor.org/stable/2397741.

[5] Nakahara, M., "Geometry, Topology and Physics", Taylor & Francis, Graduate student series in physics (2003)

[6] Chern, Shiing-Shen, "Characteristic classes of Hermitian Manifolds", Annals of Mathematics, Second Series, The Annals of Mathematics, Vol. 47, No. 1, 47 (1): 85–121 (1946), https://doi.org/10.2307/1969037

[7] C. L. Kane and E. J. Mele, "Quantum spin hall effect in graphene", Phys. Rev. Lett., 95:226801, (2005) https://link.aps.org/doi/10.1103/PhysRevLett.95.226801

[8] Kitaev, Alexei, "Periodic table for topological insulators and superconductors", AIP Conference Proceedings 1134, 22 (2009) https://doi.org/10.1063/1.3149495

[9] M. R. Zirnbauer, "Riemannian symmetric superspaces and their origin in random-matrix theory", J. Math. Phys. 37, 4986 (1996) https://doi.org/10.1063/1.531675

[10] A. Altland and M. R. Zirnbauer, "Nonstandard symmetry classes in mesoscopic normal-superconducting hybrid structures" Phys. Rev. B 55, 1142 (1997). https://doi.org/10.1103/PhysRevB.55.1142

[11] In general literature on geometric phases (like the Berry phase) $\boldsymbol{R}$ is the most common symbol for the parameter, but in this context we will use $\boldsymbol{k}$.

[12] Two maps f, g are defined as *homotopic* if there exists a continuous interpolation, which is a continuous map between the same spaces as f/g with an extra parameter $\alpha \in [0,1]$ such that setting $\alpha = 0$ produces f and setting $\alpha = 1$ produces g. Two spaces A, B of maps are *homotopy equivalent* if there exist functions from $A \to B$ and $B \to A$ which are inverses up to homotopy. In this case, we have argued that each band structure can be interpolated to another one which is equivalent to a map $\mathrm{BZ} \to \mathbf{Gr}(k,n)$, which is sufficient to prove homotopy equivalence.