

# Local formula for the $\mathbb{Z}_2$ invariant of topological insulators

Zhi Li and Roger S. K. Mong

*Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260, USA  
and Pittsburgh Quantum Institute, Pittsburgh, Pennsylvania 15260, USA*



(Received 31 May 2019; published 4 November 2019; corrected 28 February 2020)

We proposed a formula for the  $\mathbb{Z}_2$  invariant for topological insulators, which remains valid without translational invariance. Our formula is a local expression, in the sense that the contributions mainly come from quantities near a point. Using almost commute matrices, we proposed a method to approximate this invariant with local information. The validity of the formula and the approximation method is proved.

DOI: [10.1103/PhysRevB.100.205101](https://doi.org/10.1103/PhysRevB.100.205101)

## I. INTRODUCTION

One of the most important progresses of condensed matter physics in recent years is the realization of many topological phases of matter beyond the Landau-Ginzburg paradigm. While the general classification of topological phases is still in progress, the classification for gapped noninteracting fermions is well established [1–4] and shows beautiful connections to  $K$ -theory and symmetric spaces. According to the action of several discrete symmetries, systems are classified into 10 classes. In each class, systems are labeled by a topological invariant valued in  $\mathbb{Z}$  or  $\mathbb{Z}_2$ . The pattern appearing for various dimensions can be naturally explained by the Bott periodicity [5,6] and can be arranged into a periodic table.

A topological insulator, first proposed by Kane and Mele in Ref. [7], is a nontrivial system in two dimensions (2D) with time-reversal symmetry which squared to  $-1$  (AII class in the Altland-Zirnbauer classification [1]). It is characterized by the gapless helical edge modes protected by the time-reversal symmetry [8], and the band crossing in the language of topological band theory. The topological invariant in this case is a  $\mathbb{Z}_2$  number which we call Kane-Mele invariant.

For systems with translational invariance, one can get analytical formulas for the topological invariants by working in momentum space and considering essentially some vector bundles (with symmetries) [2–4] over the Brillouin zone. For example, see Refs. [7,9–13] for various formulas for the  $\mathbb{Z}_2$  Kane-Mele invariant.

While the classification is believed to be robust against disorder [3,14,15], analytical formulas are more difficult to find. Nevertheless, one can still get useful results from noncommutative geometry/topology considerations [16–19], which may manifest itself as a (Fredholm, mod 2 Fredholm, Bott, etc.) index [20–29] (although some of them are abstract definitions and do not tell us how to calculate them efficiently), or from physical considerations such as scattering theory [30]. A nice example is the following formula [31,32] for two-dimensional Chern insulators (class A):

$$\nu(P) = 12\pi i \sum_{j \in A} \sum_{k \in B} \sum_{l \in C} (P_{jk} P_{kl} P_{lj} - P_{jl} P_{lk} P_{kj}), \quad (1)$$

where  $P$  is the orthogonal projection operator onto filled bands, or equivalently the ground-state correlation matrix.  $\{A, B, C\}$  is a partition of the plane into three parts, as in Fig. 1(b). This formula reveals the local nature of the Chern number: assuming  $P_{ij}$  decays fast enough as  $|i - j| \rightarrow \infty$ , then  $\nu(P)$  can be well approximated by only summing over  $j, k, l$  near the intersection point. For example, truncate the plane with a circle as in Fig. 1(c), then the same summation (with  $A, B$ , and  $C$  now finite) provides a good estimation.

In this paper, we propose a formula for the  $\mathbb{Z}_2$  invariant for topological insulators in two dimensions, which remains valid with disorder. Importantly, our approach is *purely topological*, in the sense that we discard many geometrical information/choices such as distances and angles [see Eq. (16)]. Moreover, we only require a mobility gap instead of a spectral gap. Similar to Eq. (1), the input of our formula is the projection  $P$ . Also similar to Eq. (1), our formula is essentially a local expression, in the sense that the contribution mainly comes from quantities near a point. As a result, one can expect to calculate it with sufficient precision by a truncation near that point.

This paper is organized as follows. In Sec. II, we explain the physics intuition and give a physical derivation of our formula. In Sec. III, we formally state our formula and show that it is well defined. In Sec. IV, based on the theory of almost commuting matrices, we introduce a method to numerically calculate the invariant from a finite-size system. We present some numerical results in Sec. V. In Sec. VI, we investigate the properties of our formula and sketch the proof of our main proposition. To keep the paper more accessible, some technical details are gathered into the Appendix.

## II. INTUITION-FLUX INSERTION AND TOPOLOGICAL INVARIANT

In this section, we put Chern insulator/topological insulator on a punctured plane and insert fluxes at the origin [see Fig. 1(a)]. We will explain how the physics of flux insertion is related to topological invariant. This section aims to explain our intuition and provide a physical derivation of our formula, hence, some statements here may not very rigorous. We will establish our results carefully in the following sections.



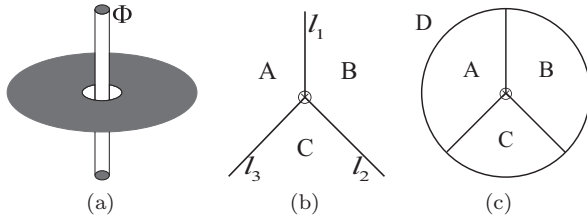


FIG. 1. (a) Insert a flux in the hole. (b) Divide the plane into three regions A, B, C. The intersection point is where a flux will be inserted. (c) Truncate the plane with a circle. Denote the region outside the circle by D. The intersection point is where a flux will be inserted.

Recall the simple case, Chern insulators, which can be realized in integer anomalous quantum Hall systems. In this case, we have the well-known Thouless charge pump [33]: when a flux unit is adiabatically inserted, it induces an annular electric field, which in turn produces a radial electric current due to the Hall effect. As a result, electrons are pushed away from (or close to, depending on the sign of the current) the origin for “one unit.” In Fig. 2 we draw the band structure for boundary states (near the puncture). Diagrammatically, when a flux unit is inserted, every occupied state moves toward top right to a lower level.

The new many-body state is not the ground state because there is a filled state above the Fermi level. Compared to the ground states, we can see that the ground state has one less electron ( $k_0$  electron in Fig. 2) than the old one (note that we are doing  $\infty - \infty$ , see comments below). The difference of number of electrons in ground states is exactly the Chern number. This is the idea behind Ref. [20]:

$$\begin{aligned} \text{Chern number} &= \text{Ind}(P, P') \\ &= \dim \text{Ker}(P - P' - 1) - \dim \text{Ker}(P - P' + 1), \end{aligned} \quad (2)$$

where  $P/P'$  is the projection operator onto filled states before/after the flux insertion,  $\text{Ind}$  is the relative index for a pair of projections, which intuitively counts the difference of their ranks (dimension of eigenvalue 1 subspace, number of filled levels in physics). Since the rank is just  $\text{Tr}(P)$  and  $\text{Tr}(P')$ , one may expect

$$\text{Ind}(P, P') \sim \text{Tr}(P - P'). \quad (3)$$

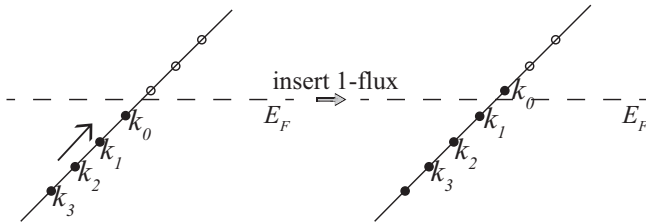


FIG. 2. Band for boundary states of a Chern insulator. • means filled, ○ means empty. After a unit-flux insertion, every filled state moves toward top right to the next level. In this process, the label  $k_i$  is tight to the electron, not the level.

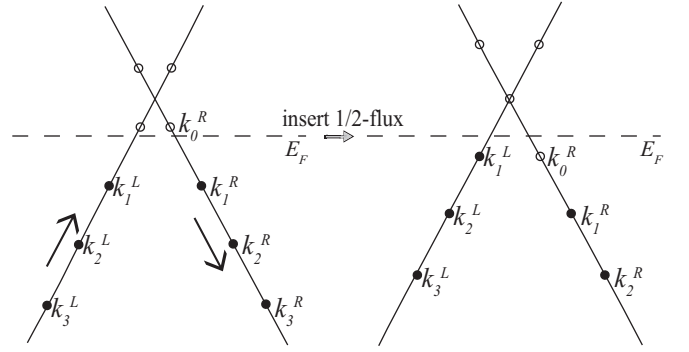


FIG. 3. Band for boundary states of a topological insulator. After a half-flux insertion, we get one more state under the Fermi level, which is geometrically near the vertex (flux).

This formula is indeed correct if  $\text{Tr}(P - P')$  is well defined, if  $(P - P')$  is trace class [34]. This is not the case for nontrivial Chern insulator though:  $\text{Ind}(P, P')$  is still well defined [20], but one needs a more complicated formula to evaluate it, which is essentially Eq. (1).

Now, we turn to topological insulators. In this case, we adiabatically insert a  $\frac{1}{2}$ -flux quanta. As shown in Fig. 3, what happens is as follows: energies for left movers increase, while energies for right movers decrease. If we assume (without loss of generality) the Fermi level is right below an empty state, the ground state after the flux insertion will have one more electron than before. We want to count the number of extra electrons to determine the Kane-Mele invariant  $\nu_{\text{KM}}(\text{mod}2)$ .

To do this, we first count the number of electrons in a finite disk with radius  $r$  (the system is still on an infinite plane, we just draw a virtual circle to define a disk). Due to time-reversal symmetry, topological insulator have zero total Hall conductance, so the number of electrons inside the disk remains unchanged under adiabatic flux insertion. However, there is a vertex state ( $k_0^R$  in Fig. 3) that is left empty, so in the new ground state the number of electrons in disk is increased by 1:

$$\Delta \langle \text{No. electrons in a (large) disk} \rangle = 1 = \nu_{\text{KM}}(\text{mod}2), \quad (4)$$

where  $\langle \dots \rangle$  means ground-state expectation value (note again that ground states before and after the flux insertion are different).

Since  $P$  is the projection onto filled states,  $H_0 = 1 - P$  can be regarded as a spectral-flattened Hamiltonian (filled = 0, empty = 1). Denote  $H_{\frac{1}{2}}$  to be the Hamiltonian after flux insertion, consider  $Q = 1 - H_{\frac{1}{2}}$  and corresponding projection matrix  $\bar{Q}$  (see Sec. III for details). We have

$$\begin{aligned} \langle N_r \rangle (\text{before}) &= \left\langle \sum_{|i| < r} a_i^\dagger a_i \right\rangle = \text{Tr}(P_r), \\ \langle N_r \rangle (\text{after}) &= \text{Tr}(\bar{Q}_r), \end{aligned} \quad (5)$$

$$\Delta \langle N_r \rangle = \text{Tr} \bar{Q}_r - \text{Tr} P_r = \text{Tr}(\bar{Q}_r - Q_r).$$

Here,  $N_r$  is the number of electrons in disk  $r$ ,  $P_r, \bar{Q}_r$  is the truncation of  $P, \bar{Q}$  [ $\bar{Q}_r$  means  $(\bar{Q})_r$ : spectral flatten before truncation]. The last equation is because  $P$  and  $Q$  have the



same diagonal elements (see Sec. III) and they are finite matrices.

Thus, we expect

$$\begin{aligned} \nu_{\text{KM}} &= \lim_{r \rightarrow \infty} \Delta \langle N_r \rangle \\ &= \lim_{r \rightarrow \infty} \text{Tr}(\bar{Q}_r - Q_r) \\ &\sim \text{Tr}(\bar{Q} - Q) \pmod{2}. \end{aligned} \quad (6)$$

One may want to apply the same idea to Chern insulator. This will just lead to  $0 = 0$ . Indeed, we still have

$$\Delta \langle \text{No. electrons in a (large) disk} \rangle = \lim_{r \rightarrow \infty} \text{Tr}(\bar{Q}_r - Q_r). \quad (7)$$

However, there will always be an electron go into (or out of) the disk adiabatically, which compensates the lost (or extra) state, so  $\Delta \langle \text{No. electrons in disk} \rangle$  in the left-hand side is always 0 in this case. This can also be seen from the (large) gauge equivalence between the two systems before and after a unit-flux insertion. For the right-hand side, since  $Q = P'$  in this case is already a projection,  $\bar{Q} = Q$ , so the right-hand side is 0. The difference between topological insulators and Chern insulators is as follows: in the former case the number of electrons go through the boundary  $r$  is 0 in average (because of zero Hall conductance), while in the latter it is nonzero and is essentially the Chern number.

As a side note, one may also consider the insertion of a unit flux and consider the difference between two ground states. A direct application of the relative index (2) gives 0. However, one may note that two terms (dimension of the kernel) in Eq. (2) come from left movers and right movers separately and one can therefore define the  $\mathbb{Z}_2$  index with one kernel. This is the idea behind Ref. [27].

### III. FORMULA FOR INFINITE SYSTEM

In this section, we will carefully define the quantities in our main formula (6) and show its well definedness. The input of our formula will be the single-body projection operator  $P$ , which is related to the spectrum-flattened Hamiltonian  $H_0 = 1 - P$ . For gapped states,  $P$  decays at least exponentially [35]:

$$P_{\mathbf{x},\mathbf{y}} < C_1 e^{-C_2 |\mathbf{x}-\mathbf{y}|}. \quad (8)$$

According to the Peierls substitution [36], if we insert a  $\frac{1}{2}$ -flux at a vertex, the new (single-body) Hamiltonian can be written as  $H_{\frac{1}{2}} = 1 - Q$ , where

$$Q_{\mathbf{x},\mathbf{y}} = s_{\mathbf{x},\mathbf{y}} P_{\mathbf{x},\mathbf{y}}. \quad (9)$$

Here,  $s_{\mathbf{x},\mathbf{y}}$  are phases such that for any loop  $l = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n = \mathbf{x}_1)$ , we have

$$s_l \stackrel{\text{def}}{=} \prod_{i=1}^n s_{\mathbf{x}_i, \mathbf{x}_{i+1}} = \begin{cases} -1, & \text{if the vertex is in the loop} \\ 1, & \text{otherwise.} \end{cases} \quad (10)$$

These phases can be chosen as follows: we divide the plane into three regions, as in Fig. 1(b).

Let  $n_{\mathbf{x},\mathbf{y}}$  be the number intersections of the straight-line segment  $(\mathbf{x}, \mathbf{y})$  with three boundaries  $l_1, l_2, l_3$ , set

$$s_{\mathbf{x},\mathbf{y}} = (-1)^{n_{\mathbf{x},\mathbf{y}}}. \quad (11)$$

We call this gauge “insert half-fluxes along the boundaries.” While  $P$  is a projection,  $Q$  no longer is. Actually, we have

$$\begin{aligned} (Q^2 - Q)_{\mathbf{x},\mathbf{y}} &= \sum_{\mathbf{z}} s_{\mathbf{x},\mathbf{z}} s_{\mathbf{z},\mathbf{y}} P_{\mathbf{x},\mathbf{z}} P_{\mathbf{z},\mathbf{y}} - s_{\mathbf{x},\mathbf{y}} P_{\mathbf{x},\mathbf{y}} \\ &= -2 s_{\mathbf{x},\mathbf{y}} \sum_{\mathbf{z}} ' P_{\mathbf{x},\mathbf{z}} P_{\mathbf{z},\mathbf{y}}, \end{aligned} \quad (12)$$

where  $\sum'$  means sum under constraint  $s_{\mathbf{x}\mathbf{y}\mathbf{z}} = -1$ . Denote  $V = Q^2 - Q$ . Since matrix elements of  $P$  decay exponentially,  $V$  is mainly supported around the vertex (hence the notation  $V$ ) due to the constraint. To be specific, we have the following:

*Proposition 1.*  $\exists C'_1, C'_2$ , such that  $|V_{\mathbf{x},\mathbf{y}}| < C'_1 e^{-C'_2 r}$  where  $r = \max\{|\mathbf{x}|, |\mathbf{y}|\}$ .

*Proof.* Let us calculate  $V_{\mathbf{x},\mathbf{y}}$ :

$$|V_{\mathbf{x},\mathbf{y}}| = \left| 2 \sum_{\mathbf{z}} ' P_{\mathbf{x},\mathbf{z}} P_{\mathbf{z},\mathbf{y}} \right| < 2C_1^2 \sum_{\mathbf{z}} ' e^{-C_2(|\mathbf{x}-\mathbf{z}|+|\mathbf{z}-\mathbf{y}|)}. \quad (13)$$

From geometry, it is obvious that  $|\mathbf{x}-\mathbf{z}| + |\mathbf{z}-\mathbf{y}| > r$  if  $s_{\mathbf{x}\mathbf{y}\mathbf{z}} = -1$ , so the summation can be controlled by

$$2C_1^2 e^{-\frac{C_2}{2}r} \sum_{\mathbf{z}} e^{-\frac{C_2}{2}(|\mathbf{x}-\mathbf{z}|+|\mathbf{z}-\mathbf{y}|)} < C'_1 e^{-C'_2 r}. \quad (14)$$

(This is a pretty crude estimation but is enough.) ■

In the following, we will call the property in this proposition as “exponential decay property” (EDP). Intuitively,  $Q^2 - Q$  satisfies EDP means the deviation of  $Q$  from a projection mainly comes from states near the vertex point. If we spectral flatten  $Q$  to  $\bar{Q}$  (for eigenvalues  $\lambda \leq \frac{1}{2}$ , convert it to 0, otherwise convert it to 1), we anticipate that  $Q - \bar{Q}$  is mainly supported near the vertex. Actually,  $Q - \bar{Q}$  also obeys EDP, but we do not need this result. We only need the following:

*Proposition 2.*  $Q - \bar{Q}$  is trace class.

*Proof.*  $|x - \bar{x}| \leq 2|x^2 - x|$  for  $\forall x \in \mathbb{R}$ , so  $|Q - \bar{Q}| \leq 2|Q^2 - Q| = 2|V|$  as an operator (note that they commute). Since  $V$  obeys EDP,  $V$  must be trace class (see the corollary after Lemma 2 in the Appendix, Sec. A 1), so is  $Q - \bar{Q}$ . ■

Therefore, it is legal to define a “trace over vertex states” as

$$\text{Tr}^v(Q) = \text{Tr}(Q - \bar{Q}). \quad (15)$$

Note that in the definition of  $\bar{Q}$  we can arbitrarily choose the chemical potential  $\mu \in (0, 1)$ , so the  $\text{Tr}^v(Q)$  should be naturally understood as mod 1. In the case of topological insulator,  $Q$  has time-reversal symmetry, every state is Kramers paired, so  $\text{Tr}^v(Q)$  can be naturally understood as mod 2. We will see in the following that it is  $\text{Tr}^v(Q) \pmod{2}$  [instead of  $\text{Tr}^v(Q)$  itself for a fixed “chemical potential”] that has good properties. Also note that  $Q$  is not trace class in general, so we cannot define  $\text{Tr}^v(Q)$  as  $\text{Tr}(Q) \pmod{2}$ .

According to the above analysis, this expression is well defined and the contributions mainly come from states near the vertex; it is a local expression. Interestingly, this local expression turns out to be independent of the flux-insertion point we choose; it only depends on the state itself. Moreover, it is an integer and is topologically invariant. Our main proposition is as follows:

*Main proposition.*  $\text{Tr}^v(Q)$  equals to the Kane-Mele invariant.



The derivation of our proposition is in Sec. VI and the Appendix. Before going on, we give three comments on our formula.

(1) There is another construction of  $Q$ , closely related to the one given by Eq. (10):

$$Q = \begin{bmatrix} AA & -AB & -AC \\ -BA & BB & -BC \\ -CA & -CB & CC \end{bmatrix}, \quad (16)$$

where  $AB$  means  $P_{AB}$ , a block in the original matrix  $P$ . If we consider a circle with many sites on it, it still gives us a total phase  $-1$ . In this case,

$$Q^2 - Q = 2 \begin{bmatrix} 0 & ACB & ABC \\ BCA & 0 & CAB \\ CBA & BAC & 0 \end{bmatrix}, \quad (17)$$

where  $ACB$  means  $P_{AC}P_{CB}$ , etc. It is still concentrated near the vertex (satisfies EDP), as long as the partition is good.<sup>1</sup> So, we can follow the same procedure to define a new  $\text{Tr}^v(Q)$ .

The  $Q$  defined here is *not* unitary equivalent to the one in Eq. (10): they have different spectra in general. However, in Sec. VI we will show that  $\text{Tr}^v(Q)$  defined from them are equal (mod 2). We call the  $Q$  in Eq. (16) the topological one, denoted by  $Q^t$ , because its definition does not depend on the geometric information such as “straight-line segments.” The  $Q$  in Eq. (10) will be called the geometric one, denoted by  $Q^g$ . It has the advantage of gauge invariance and many quantities [like  $\text{spec}(Q)$ ] defined from it are manifestly independent of the partition.

(2) For systems in the DIII class ( $\text{TRS}^2 = -1$ ,  $\text{PHS}^2 = 1$  where TRS is time-reversal symmetry and PHS is particle-hole symmetry), our formula can be simplified. Indeed, since the original system has PHS,

$$K_{ph}^\dagger (2P - 1) K_{ph} = 2P - 1, \quad (18)$$

where  $(2P - 1)$  is the spectral-flattened Hamiltonian with spectrum  $= \{\pm 1\}$ .  $K_{ph}$  is an onsite action, and commutes with the operation from  $P$  to  $Q$ , so the same equation holds for  $Q$ . Therefore, the spectra of  $Q$  is symmetric with respect to  $\frac{1}{2}$ :

$$\sigma(Q) = 1 - \sigma(Q). \quad (19)$$

Now, for a spectrum  $q$  such that  $q \neq \frac{1}{2}$ , the Kramers degeneracy and PHS provides us a fourfold  $\{q, q, 1 - q, 1 - q\}$ , which contributes 0 to  $\text{Tr}^v(Q) \pmod{2}$ . So,

$$v = \text{Tr}^v Q = \{\text{No. Kramers pairs at } \frac{1}{2}\} \pmod{2}. \quad (20)$$

(3) The input  $P$  is the correlation matrix for an infinite system. If we start with a finite system, say a topological insulator on the sphere, then our formula always gives 0. Mathematically, this is because both  $\text{Tr}(\bar{Q})$  and  $\text{Tr}(Q) = \text{Tr}(P)$  are even due to time-reversal symmetry. Physically, it is because when inserting a flux at some point, it is unavoidable to insert another flux at somewhere else for a closed geometry, then our formula counts the vertex states at both points. To

get the right invariant, we need to “isolate” the physics at one vertex.

#### IV. APPROXIMATION FROM FINITE SYSTEM

Although the input of our formula is an infinite-dimensional operator  $P$ , our formula is a trace of vertex states, which should only depend on the physics near the origin. Let us truncate the plane with a circle  $r$  [see Fig. 1(c)]. Denote  $P_N$  and  $Q_N$  to be the truncation of  $P$  and  $Q$ , where  $N$  is the number of sites inside the circle  $N \sim r^2$ . We expect that one can approximate the invariant with data near the origin, i.e., with matrix elements of  $P_N$  or equivalently  $Q_N$ .

However, a naive limit  $\lim_{N \rightarrow \infty} \text{Tr}(Q_N - \bar{Q}_N)$  is wrong: it will give  $\lim \text{Tr}(P_N) \pmod{2}$  since  $\text{Tr}(\bar{Q}_N)$  is even. This is because  $\bar{Q}_N \neq (\bar{Q})_N$ . Physically,  $Q_N$  and  $Q$  do have similar “vertex states.” However, unlike  $Q$ ,  $Q_N$  also includes boundary contributions [see Eq. (22)], which need to be excluded.

We claim that we can use the following algorithm to approximate our invariant.

(i) Construct a matrix  $V_N$  by

$$V_N = -2s_{x,y} \sum_{\substack{\mathbf{z} \in (ABC) \\ s_{xyz} = -1}} P_{x,z} P_{z,y}. \quad (21)$$

$V_N$  will be almost commuted with  $Q_N$  and it will tell us whether a state is near the vertex or the boundary.

(ii) Find approximations  $Q'_N, V'_N$  for  $Q_N, V_N$  so that they indeed commute.<sup>2</sup>

(iii) Simultaneously diagonalize  $Q'_N$  and  $V'_N$  to get pairs of eigenvalues  $(q', v')$ . Sum over all the eigenvalues  $q'$  such that  $v' \neq 0$ .

The summation will converge to  $\text{Tr}^v(Q)$  as  $N \rightarrow \infty$ .

In the following we explain the algorithm in detail. First of all, we have

$$\begin{aligned} (Q_N^2 - Q_N)_{x,y} &= -s_{x,y} \left[ 2 \sum_{\substack{\mathbf{z} \in ABC \\ s_{xyz} = -1}} P_{x,z} P_{z,y} + \sum_{\mathbf{z} \in D} P_{x,z} P_{z,y} \right] \\ &\stackrel{\text{def}}{=} (V_N)_{x,y} + (W_N)_{x,y}. \end{aligned} \quad (22)$$

Here,  $V_N$  is supported near the center, while  $W_N$  is supported near the boundary. This means the deviation of  $Q_N$  to a projection happens both near the vertex and the boundary.

We can also work in the topological construction of  $Q$ . In this case,

$$\begin{aligned} Q_N^2 - Q_N &= 2 \begin{bmatrix} 0 & ACB & ABC \\ BCA & 0 & CAB \\ CBA & BAC & 0 \end{bmatrix} - \begin{bmatrix} ADA & ADB & ADC \\ BDA & BDB & BDC \\ CDA & CDB & CDC \end{bmatrix} \\ &= V_N + W_N. \end{aligned} \quad (23)$$

<sup>1</sup>For example, the one in Fig. 1(b) is good. However, if we rotate  $l_2$  toward  $l_1$  and deform it a little bit so they are parallel at infinity, then  $Q^2 - Q$  does not satisfy EDP and a convergence problem will occur.

<sup>2</sup>In practice, there is some arbitrariness to find  $Q'_N, V'_N$ . What we do is a joint approximation diagonalization (JAD) and then make them commute according to some rules. For example, one may make all  $v'$  such that  $|v'| > \epsilon$  to zero. Another rule is indicated in Sec. V.



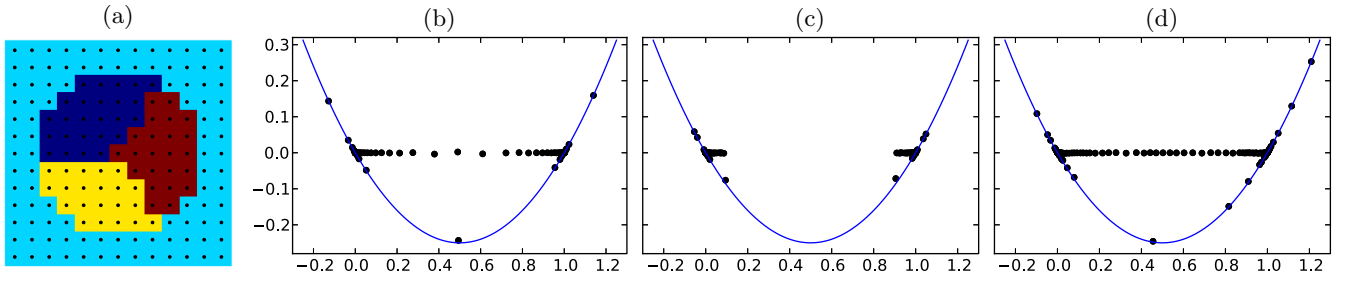


FIG. 4. Numerical results. (a) The geometry for computing the topological invariants. The regions A, B, and C are represented by colors yellow, red, and navy blue, respectively. (Region D is represented by cyan.) (b), (c) Numerical results for the Hamiltonian (30) with (b)  $m = 1.6$  and (c)  $m = 2.4$ . Shown in the plot are eigenvalues of  $V'_N$  vs  $Q'_N$ . These results are generic; dots along the  $x$  axis represent boundary states; dots near (0,0) and (1,0) are bulk states; dots on the parabola represent vertex states. (d) Numerical result from the Hamiltonian (31), which strongly breaks particle-hole symmetry.

In both constructions,  $Q_N, V_N, W_N$  should almost commute, and  $V_N, W_N$  are almost orthogonal since they are mainly supported in different regions (“almost” means relevant expressions approach 0 as  $N \rightarrow \infty$ ).

*Proposition 3.* (1)  $Q_N, V_N, W_N$  defined above satisfies

$$\begin{aligned} \| [Q_N, V_N] \| &< \epsilon, & \| [Q_N, W_N] \| &< \epsilon, \\ \| V_N W_N \| &< \epsilon, & \| W_N V_N \| &< \epsilon, \end{aligned} \quad (24)$$

where the norm  $\| \cdot \|$  is the  $L^2$  norm (maximal singular value),  $\epsilon \sim p(r)e^{-r}$  where  $p(r)$  is a polynomial of  $r$ . (2) There exist Hermitian matrices  $Q'_N, V'_N, W'_N$  as approximations of  $Q_N, V_N, W_N$  in the sense that

$$\| Q_N - Q'_N \| < \rho^2, \quad \| V_N - V'_N \| < \rho, \quad \| W_N - W'_N \| < \rho, \quad (25)$$

such that

$$[Q'_N, V'_N] = [Q'_N, W'_N] = V'_N W'_N = W'_N V'_N = 0, \quad (26)$$

$$Q_N^2 - Q'_N = V'_N + W'_N. \quad (27)$$

Here,  $\rho$  can be chosen as  $F(\epsilon)\epsilon^{1/10}$  (independent of  $N$ ) where the function  $F(x)$  grows slower than any power of  $x$ .

*Proof.* (1) Straightforward calculation. (2) It is easy to check  $\|Q_N\|$  and  $\|V_N\|$  are finite, independent of  $N$  [one way to do this is to prove it for the topological construction  $Q$  in Eq. (16) and use the relationship between two constructions as in Property 2]. According to Lin’s theorem [37],  $\exists Q'_N, V'_N$  such that  $\|Q_N - Q'_N\|, \|V_N - V'_N\| < \delta$ , and  $[Q'_N, V'_N] = 0$ . Moreover [38], we can choose  $\delta = E(1/\epsilon)\epsilon^{1/5}$  where the function  $E(x)$  grows slower than any power of  $x$ , independent of  $N$ .

Define  $W'_N = Q_N^2 - Q'_N - V'_N$ , then  $W'_N, Q'_N, V'_N$  can be simultaneously diagonalized. Since  $W_N = Q_N^2 - Q'_N - V_N$ ,  $\|Q_N - Q'_N\|, \|V_N - V'_N\| < \delta$ , so we have  $\|W_N - W'_N\| \lesssim \delta$  and

$$\| V'_N W'_N \| = \| (V'_N - V_N + V_N)(W'_N - W_N + W_N) \| \lesssim \epsilon + \delta \sim \delta. \quad (28)$$

This means for each pair of eigenvalues  $(v'_N, w'_N)$ , at least one of them should be smaller than  $\sqrt{\delta}$ . We manually make these eigenvalues to be 0, while fixing  $v'_N + w'_N$ .

The new  $V'_N$  and  $W'_N$  would be strictly orthogonal, and still commute with  $Q'_N$ , and still obey  $Q_N^2 - Q'_N = V'_N + W'_N$ . Moreover, now  $\|V_N - V'_N\| \sim \delta + \sqrt{\delta} \sim \sqrt{\delta} \stackrel{\text{def}}{=} \rho$ . ■

Having  $Q'_N, V'_N, W'_N$  exactly commute, and  $V'_N, W'_N$  exactly orthogonal, we use them to distinguish vertex contributions and boundary contributions. We simultaneously diagonalize them and get triples  $(q', v', w')$ . Different contributions are then identified as follows (the reason for this identification is evident) (see Fig. 4):

- (i)  $v' \neq 0, w' = 0$ : vertex states;
- (ii)  $v' = 0, w' \neq 0$ : boundary states;
- (iii)  $q' = 0$  or  $1, v' = w' = 0$ : bulk states.

We anticipate that the summation of  $q'$  over vertex states will be a good approximation of  $\text{Tr}^v(Q)$ .

*Proposition 4 (Finite-size approximation).* After the above procedure,

$$\lim_{N \rightarrow \infty} \sum_{\text{vertex states}} q' = \text{Tr}^v(Q). \quad (29)$$

The proof is in the Appendix, Sec. A 2.

## V. NUMERICAL RESULTS

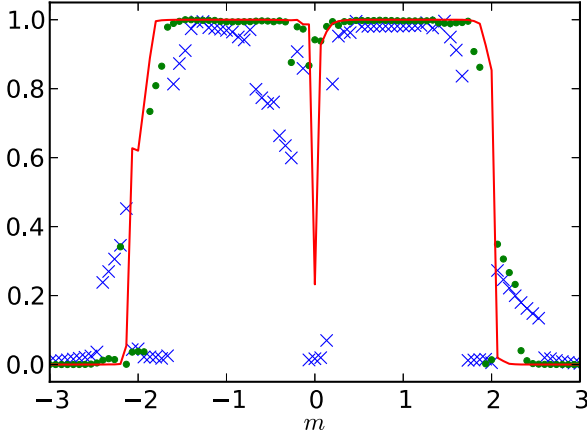
For our numerical results we use the Bernevig-Hughes-Zhang (BHZ) model [39] on a square lattice with Rashba coupling and scalar/valley disorder:

$$\begin{aligned} H &= \int_{\mathbf{k}} (H_0 + H_R) d^2 \mathbf{k} + \sum_{\mathbf{r}} H_D(\mathbf{r}), \\ H_0(\mathbf{k}) &= v(\tau^x \sigma^z \sin k_x + \tau^y \sin k_y) \\ &\quad + (m - t \cos k_x - t \cos k_y) \tau^z, \\ H_R(\mathbf{k}) &= r(\sigma^x \sin k_y - \sigma^y \sin k_x), \\ H_D(\mathbf{r}) &= V(\mathbf{r}, +) \frac{1 + \tau^z}{2} + V(\mathbf{r}, -) \frac{1 - \tau^z}{2}. \end{aligned} \quad (30)$$

Here there are four degrees of freedom per site, with  $\tau$  and  $\sigma$  acting on valley and spin space, respectively.

In Figs. 4(b) and 4(c), we show the computation of the topological invariant of this model for  $v = t = 1, r = \frac{1}{2}$ , with the disorder  $V(\mathbf{r}, \pm)$  sampled uniformly from the interval  $[-0.4, 0.4]$ . Figure 4(b) shows a topological phase with  $m = 1.6$ , while Fig. 4(c) shows a trivial phase with  $m = 2.4$ .





	total system size	radius of ABC region
blue crosses	$7 \times 7$	2.4
green dots	$13 \times 13$	4.5
red line	$25 \times 25$	8.9

FIG. 5. The sum  $\text{Tr}^v Q$  (approximating the invariant  $\nu$ ) of a finite system for various  $m$ . The system sizes are shown in the table. [For example, the green crosses show data computed with Fig. 4(a)'s geometry.]

These plots show the eigenvalues  $(q', v')$  of the matrices  $Q'_N$  and  $V'_N$ , respectively ( $q'$  is along the  $x$  axis and  $v'$  is along  $y$  axis). Recall that we have  $(q')^2 - q' = v' + w'$  and  $v'w' \approx 0$ , hence points  $(q', v')$  either live on the parabola  $y = x^2 - x$  (if  $v' \neq 0$ ) or along the  $x$  axis (if  $v' = 0$ ). According to our analysis, points along the  $x$  axis represent boundary states; points near  $(0,0)$  and  $(1,0)$  represent bulk states; all other points on the parabola correspond to vertex states. As a comparison, Fig. 4(c) shows the trivial region where there are (mostly) only have bulk states. The goal of the numerical procedure outlined in the previous section is to isolate the vertex states which live near the intersection of A, B, and C.

As the model (30) (without disorder) is particle-hole symmetric, the resulting eigenvalues are reflection symmetric ( $q' \rightarrow 1 - q'$ ). Disorder only breaks this symmetry very weakly. To break this mirror symmetry, we construct a spinful model with three valleys,

$$\begin{aligned}
 H_0 = & 0.3\lambda^3 + 0.4\lambda^8 \\
 & + (0.5\lambda^1 + 0.6\lambda^4)\sigma^z \sin k_x + (1.2\lambda^2 - 0.3\lambda^5) \sin k_y \\
 & + (0.5\lambda^1 - 0.7\lambda^3 - 0.3\lambda^4 + 1.1\lambda^6 - 0.6\lambda^8) \cos k_x \\
 & + (0.5\lambda^1 - 0.7\lambda^3 - 0.4\lambda^4 + 1.0\lambda^6 - 0.6\lambda^8) \cos k_y,
 \end{aligned} \tag{31}$$

where  $\lambda^1, \dots, \lambda^8$  are the Gell-Mann matrices acting on valley space. We retain the Rashba term with  $r = 0.1$ , and disorder (for the three valleys independently) sampled from  $[-0.4, 0.4]$ . The spectrum  $(q', v')$  is shown in Fig. 4(d), with  $\nu$  evaluating to 1.02 indicating a QSH phase.

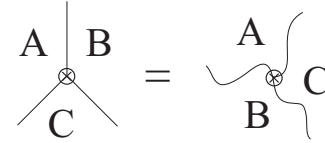
In Fig. 5, we plot the result of our formula (29) for model (30) as a function of  $m$ . (The data are computed for a single disorder realization.) For the computation of  $\text{Tr}^v Q$ , we distinguished the vertex states (from the bulk and edge

states) by only considering points satisfying  $q' < 0$ ,  $q' > 1$ , or  $v' < \frac{(q')^2 - q'}{2}$ . We see that the system is in the quantum spin Hall (QSH) phase for the bulk of  $-2 \lesssim m \lesssim 2$ . As the Hamiltonian  $H_0 + H_R$  is gapless (Dirac-type) at  $m = 0$ , we expect a small sliver of metallic phase near  $m \approx 0$  from disorder. (In general, the metallic phase is stable in the AII class, hence, we do not expect direct transition between the QSH and trivial phases.) We see that as the system size increases (so that it is large compared with the correlation length  $\xi$ ), the invariant approaches 0 or 1 to the gapped phases.

## VI. PROPERTY AND PROOF

In this section, we will investigate the property of  $\text{Tr}^v(Q)$  and derive our main proposition step by step.

*Property 1 (Gauge invariance).* Fixing the position of the flux and working in the geometric definition, then  $\text{Tr}^v(Q)$  does not depend on the actual partition of the plane. For example, the following partition and the ordering of A, B, C will give the same  $\text{Tr}^v(Q)$ :



This is because different partitions correspond to different gauge choices in the Pierls substitution. Indeed, fix a reference point  $*$ , define  $u_x = s_{*,x}/s'_{*,x}$  where  $s, s'$  are the phases for two partitions. Since  $s_{*,x}s_{x,y}s_{y,*} = s_{*xy} = s'_{*,x}s'_{x,y}s'_{y,*}$ , we have

$$s'_{x,y} = u_x s'_{x,y} \bar{u}_y. \tag{32}$$

Thus,  $Q' = U Q U^\dagger$  and they have the same spectra.

*Property 2.*  $\text{Tr}^v(Q')$  defined from topological  $Q$  (for good partitions) and  $\text{Tr}^v(Q^g)$  defined from geometric  $Q$  are equal (mod 2).

*Proof.* We calculate  $Q' - Q^g$  and find that

$$(Q' - Q^g)_{x,y} = \begin{cases} -2P_{x,y}, & \mathbf{x}, \mathbf{y} \text{ belong to different regions} \\ & \text{and } (\mathbf{x}, \mathbf{y}) \text{ intersect 2 boundaries} \\ 0, & \text{otherwise.} \end{cases} \tag{33}$$

From geometry we can see if  $|\mathbf{x}| > r$ , then the first condition is satisfied only if  $|\mathbf{y} - \mathbf{x}| > O(r)$  where  $r = \max\{|\mathbf{x}|, |\mathbf{y}|\}$ . So,  $Q' - Q^g$  satisfies a EDP:  $|(Q' - Q^g)_{x,y}| < C_1 e^{-C_2 r}$  and thus is trace class. Therefore,

(i)  $\text{Tr}(Q' - Q^g) = \lim_{N \rightarrow \infty} \text{Tr}(Q'_N - Q^g_N) = 0$  since they always have the same diagonal elements (note that we need trace class condition for the first equation evolving limit [34]);

(ii)  $\overline{Q'} - \overline{Q^g} = (Q' - \overline{Q'}) - (Q^g - \overline{Q^g}) - (Q' - Q^g)$  is also trace class.

Due to time-reversal symmetry,

$$\begin{aligned}
 & \text{Tr}(\overline{Q'} - \overline{Q^g}) \\
 &= \text{Ind}(\overline{Q'}, \overline{Q^g}) \\
 &= \dim \text{Ker}(\overline{Q'} - \overline{Q^g} - 1) - \dim \text{Ker}(\overline{Q'} - \overline{Q^g} + 1) \\
 &= 0(\text{mod } 2),
 \end{aligned} \tag{34}$$



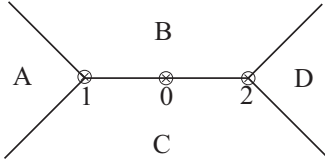


FIG. 6. Divide the plane into four regions  $A + B + C + D$ . Insert a  $\frac{1}{2}$ -flux for each vertex (1 and 2). We will show that it is “equal” to insert a unit flux at the middle (point 0).

where  $\text{Ind}(\cdot, \cdot)$  is the index for a pair of projections [20]. So, we have

$$\begin{aligned} \text{Tr}^v(Q^t) - \text{Tr}^v(Q^g) &= \text{Tr}(Q^t - \bar{Q}^t - Q^g - \bar{Q}^g) \\ &= \text{Tr}(Q^t - Q^g) - \text{Tr}(\bar{Q}^t - \bar{Q}^g) \\ &= 0(\text{mod}2). \end{aligned} \quad (35)$$

Now, we insert two  $\frac{1}{2}$ -fluxes at different positions far away from each other. We divide the plane into four regions, as in Fig. 6. Again, we “insert half-fluxes along the boundaries” and write the Hamiltonian as

$$S_{\mathbf{x},\mathbf{y}} = s_{\mathbf{x},\mathbf{y}} P_{\mathbf{x},\mathbf{y}}, \quad (36)$$

where  $s_{\mathbf{x},\mathbf{y}}$  are defined similar to Eq. (11) by the new partition. We have

$$\begin{aligned} S^2 - S &= -2s_{\mathbf{x},\mathbf{y}} \sum_{\substack{\mathbf{z} \\ s_{\mathbf{xyz}} = -1}} P_{\mathbf{x},\mathbf{z}} P_{\mathbf{z},\mathbf{y}} \\ &= -2s_{\mathbf{x},\mathbf{y}} \left( \sum_{\substack{\mathbf{z} \\ O_1 \in (\mathbf{xyz}) \\ O_2 \notin (\mathbf{xyz})}} + \sum_{\substack{\mathbf{z} \\ O_1 \notin (\mathbf{xyz}) \\ O_2 \in (\mathbf{xyz})}} \right) P_{\mathbf{x},\mathbf{z}} P_{\mathbf{z},\mathbf{y}} \\ &\stackrel{\text{def}}{=} (V_1)_{\mathbf{x},\mathbf{y}} + (V_2)_{\mathbf{x},\mathbf{y}}. \end{aligned} \quad (37)$$

Here,  $V_i$  ( $i = 1, 2$ ) is the “vertex” term for two junctions, respectively. As in Sec. III,  $V_i$  ( $i = 1, 2$ ) satisfies EDP for vertex  $i$  and  $S - \bar{S}$  is trace class, so  $\text{Tr}^v(S)$  is well defined.

As  $\text{dist}(1, 2) \rightarrow \infty$ , we have  $V_1 V_2 = V_2 V_1 \rightarrow 0$ . In the limiting case where  $V_1 V_2 = V_2 V_1 = 0$  exactly, one can simultaneously diagonalize them and at least one eigenvalue for a common eigenstate should be 0. This means each “vertex state” of  $S$  (those states with  $S \neq 0, 1$ ) is located at junction 1 or 2. Moreover, define  $Q_1$  as the matrix corresponding to a  $\frac{1}{2}$ -flux insertion at point 1 with partition  $A + B + CD$ ,  $Q_2$  corresponding to the insertion at point 2 with partition  $AB + C + D$ , then the effect of  $S$  for a state near junction  $i$  will be close to the effect of  $Q_i$ , so one anticipates that

$$\text{Tr}^v(S) \approx \text{Tr}^v(Q_1) + \text{Tr}^v(Q_2). \quad (38)$$

In the case where  $\text{dist}(1, 2)$  is large but not infinity, some vertex states of  $S$  may come from the coupling of two vertex states at different fluxes. However, it is not difficult to convince that Eq. (38) still holds. The physics here is

similar to that for the two-states systems: due to the weak but nonzero coupling (off-diagonal elements), the eigenstates will be approximately

$$|\phi \pm\rangle = \frac{1}{\sqrt{2}}(|1\rangle \pm |2\rangle). \quad (39)$$

Here, we cannot say a vertex state of  $R$  is located at one of the fluxes. However, the summation of eigenvalues for  $|\phi \pm\rangle$  is still equal to that for  $|1\rangle$  and  $|2\rangle$ .

*Property 3 (Additivity).* Under technical assumptions, as the distance between two fluxes goes to infinity, the vertex spectrum of  $S$  “decouples”:

$$\lim_{\text{dist}(1,2) \rightarrow \infty} \text{Tr}^v(S) - [\text{Tr}^v(Q_1) + \text{Tr}^v(Q_2)] = 0. \quad (40)$$

The proof is in the Appendix, Sec. A 1.

*Property 4.*  $\text{Tr}^v(S) = 0 \pmod{2}$ . This is an exact equation, no matter whether  $\text{dist}(1, 2)$  is large or not.

The idea is, if one looks from a large distance, inserting two  $\frac{1}{2}$ -fluxes is approximately equivalent to insert a 1-flux, which is (singularly) gauge equivalent to 0-flux, so that no vertex states appear in the spectrum at all.

*Proof.* We work in the AB gauge, where the vector potential of a flux satisfies

$$|\mathbf{A}(\mathbf{r})| = \frac{\text{flux}}{2\pi r}, \quad \mathbf{A}(\mathbf{r}) \perp \mathbf{r}. \quad (41)$$

We still use  $S$  to denote the Hamiltonian in this gauge. Denote  $T$  to be the Hamiltonian for the case of inserting 1-flux at the center of two half-fluxes.  $S$  and  $T$  should be close outside the center. We will prove that  $S - T$  is trace class in the Appendix, Sec. A 3. Then, similar to the proof of Property 1.2,

$$\text{Tr}(S - T) = \lim_{N \rightarrow \infty} \text{Tr}(S_N - T_N) = 0 \quad (42)$$

since they have the same diagonal elements.  $\bar{S} - T = (\bar{S} - S) + (S - T)$  is also trace class, so

$$\begin{aligned} \text{Tr}(\bar{S} - T) &= \text{Ind}(\bar{S}, T) \\ &= \dim \text{Ker}(\bar{S} - T - 1) - \dim \text{Ker}(\bar{S} - T + 1) \\ &= 0(\text{mod}2), \end{aligned} \quad (43)$$

due to time-reversal symmetry. Therefore,

$$\text{Tr}^v(S) = \text{Tr}(S - \bar{S}) = \text{Tr}(S - T) - \text{Tr}(\bar{S} - T) = 0(\text{mod}2). \quad (44)$$

*Property 5.*  $\text{Tr}^v(Q)$  is an integer(mod2) independent of the position of the flux.

*Proof.* For every pair of points 1 and 2, we have

$$\begin{aligned} \text{Tr}^v(Q_1) - \text{Tr}^v(Q_2) &= [\text{Tr}^v(Q_1) + \text{Tr}^v(Q_3)] \\ &\quad - [\text{Tr}^v(Q_2) + \text{Tr}^v(Q_3)]. \end{aligned} \quad (45)$$

Due to Properties 3 and 4, it can be arbitrarily close to 0 (mod 2) as  $\text{dist}(1, 3)$  and  $\text{dist}(2, 3)$  goes to infinity. So we must have

$$\text{Tr}^v(Q_1) = \text{Tr}^v(Q_2)(\text{mod}2). \quad (46)$$

Plug this into Eq. (40) and use again Property 4, we obtain that  $\text{Tr}^v(Q_1) = \text{Tr}^v(Q_2) \in \mathbb{Z}_2$ .

This property already shows that  $\text{Tr}^v(Q)$  is a  $\mathbb{Z}_2$  invariant for topological insulators which only depends on the state



itself. The only natural identification is the Kane-Mele invariant.

*Property 6.*  $\text{Tr}^v(Q)$  is equal to the Kane-Mele invariant.

*Proof.* Denote our invariant as  $\nu$ . For two gapped time-reversal-symmetric systems  $A$  and  $B$ , we stack them (without hopping/interaction) and denote the new system  $A \oplus B$ . Obviously,  $\nu(A \oplus B) = \nu(A) + \nu(B)$ . From the classification of topological insulator [1–3] for AII class, the Kane-Mele invariant  $\nu_{\text{KM}}$  is the only invariant with this property. It follows that

$$\nu = k\nu_{\text{KM}}, \quad (47)$$

where  $k = 0$  or  $1$ .

To prove  $k = 1$ , it is enough to verify the existence of one system with  $\nu = 1$ . To this end, consider a translational invariant system in the DIII class with nontrivial  $\mathbb{Z}_2$  invariant. In this case, according to Eq. (20), we have

$$\nu = \{\text{No. Kramers pairs at } \frac{1}{2}\} = 1 \pmod{2}. \quad (48)$$

The last equation can be obtained by considering the band structure, due to translational invariance: a Kramers pair at  $\frac{1}{2}$  correspond to the a band crossing. ■

## VII. CONCLUSION

In this paper, we propose a formula (15) for the  $\mathbb{Z}_2$  invariant for topological insulators in 2D, which remains valid with disorder. The intuition behind our formula is flux-insertion-induced spectral flow, which manifests itself as the difference of numbers of electrons in the ground states. The formula works by taking the single-body projection matrix  $P$  (or ground-state correlation function) as the input, performing a Peierls substitution (either geometrical or topological), and then take the “trace over vertex states.” Our formula is a local expression, in the sense that the contribution mainly comes from quantities near an arbitrarily but fixed point. The validity of this formula is proved indirectly by showing its properties (gauge invariance, additivity, integrality, etc.). All properties are physically explained and mathematically proved.

Due to the local property of our formula, it can be well approximated with partial knowledge of the projection matrix. In this case, we construct “vertex matrix” and “boundary matrix” which almost commute. Using an interesting parabola, the vertex contributions are separated out. The validity of this algorithm is proved and verified numerically.

Similar ideas may be used in the case of other symmetries and other dimensions. For example, for class A in 2D (Chern insulator), an infinitesimal flux insertion will reproduce Eq. (1). It would be interesting to work out other cases to see if one can get (and prove) a new formula. Last but not least, the idea may also be extended to some interacting cases. It would be interesting to explore such generalizations.

## ACKNOWLEDGMENTS

This work is supported from NSF Grant No. DMR-1848336. Z.L. is grateful to the PQI fellowship at University of Pittsburgh.

## APPENDIX: MORE TECHNICALITIES IN THE PROOF

### 1. Proof of additivity

In this Appendix, we will prove Property 3 in Sec. VI. The proof is a little bit technical, but the physics idea is simple: vertex states of  $S$  comes from those of  $Q_1$  and  $Q_2$ .

*Lemma 1 (Almost orthogonal vectors).* Now, we have  $N$  unit vectors  $u_n$  in a  $d$ -dimensional linear space s.t.  $|(u_i, u_j)| < \sigma$  for each  $i \neq j$ . If  $\sigma < \frac{1}{\sqrt{2d}}$ , then  $N < 2d - 1$ .

*Proof.* Let  $A = (A_{i,j}) = ((u_i, u_j))$  be the Gram matrix of  $\{u_i\}$ . Denote  $\lambda_1, \dots, \lambda_N$  the eigenvalues of  $A$ . Since  $u_i \in \mathbb{C}^d$ ,  $\text{rank}(A) \leq d$ , at most  $d$  of them are nonzero. By Cauchy inequality,  $\sum \lambda_i^2 \geq (\sum \lambda_i)^2/d = (\text{Tr } A)^2/d = N^2/d$ . On the other hand,  $\sum \lambda_i^2 = \text{Tr}(A^2) = \sum |(u_i, u_j)|^2 < N + N(N-1)\sigma^2$ . Therefore,

$$\frac{N^2}{d} < N + N(N-1)\sigma^2 < N + N(N-1)\frac{1}{2d}. \quad (A1)$$

Solving this inequality, we find  $N < 2d - 1$ . ■

*Lemma 2 (An estimation of the eigenvalue distribution).*

Assume a Hermitian matrix  $A$  satisfying the exponential decay property (EDP)  $|A_{i,j}| < C_1 p(t) e^{-C_2 t}$  where  $t = \max\{|i|, |j|\}$ ,  $p(t)$  is a monic (leading coefficient=1) polynomial. The number of eigenvalues outside  $(-\epsilon, \epsilon)$  is bounded by  $O(\frac{1}{C_2^2} \ln^2 \frac{C_1}{C_2^2 \epsilon})$ , where  $\alpha = 2 + \deg p$ .

*Proof.* For a unit eigenvector  $x$ ,  $Ax = ax$ ,  $|a| > \epsilon$ , we separate  $x$  into two parts  $x = y \oplus z$  according to whether the label is inside or outside a circle:  $y_i = 0$  if  $|i| > r$ ,  $z_i = 0$  if  $|i| < r$ . The radius  $r$  will be determined later (depend on  $\epsilon$ ).

We claim that  $\|z\| < \delta \stackrel{\text{def}}{=} \frac{C_1 r^{3/2} p(r) e^{-C_2 r}}{\epsilon}$ . Indeed,  $ay \oplus az = Ax$ ,  $\|x\| = 1$ . According to the Cauchy inequality we have

$$\|\epsilon z\|^2 < \|az\|^2 < \sum_{|i| > r, j} |A_{i,j}|^2 \lesssim [C_1 r^{3/2} p(r) e^{-C_2 r}]^2. \quad (A2)$$

Here,  $\lesssim$  (means inequality up to constant) can be verified by doing integral. Denote the number of eigenvalues larger than  $\epsilon$  to be  $N$ :  $Ax_n = a_n x_n$ ,  $n = 1, \dots, N$ . Without loss of generality, we can assume they are orthogonal, so

$$(x_i, x_j) = 0 \Rightarrow |(y_i, y_j)| = |(z_i, z_j)| < \delta^2. \quad (A3)$$

Now, we have  $N$  unit vectors  $u_n = \frac{y_n}{\sqrt{1-z_n^2}}$  in dimension  $d \propto r^2$

whose inner products with each other are less than  $\sigma \stackrel{\text{def}}{=} \frac{\delta^2}{1-\delta^2}$ . We choose  $r$  large enough so that

$$\sigma < \frac{1}{\sqrt{2d}} \sim \frac{1}{r}. \quad (A4)$$

According to Lemma 1,  $N < 2d = O(r^2)$ . We can solve Eq. (A4) to estimate  $r$  (thus  $N$ ). Roughly, set  $\sigma \sim (\frac{C_1 r^{3/2} p(r) e^{-C_2 r}}{\epsilon})^2 = \frac{1}{r}$ , let  $x = C_2 r$ , we find  $e^x = \frac{C_1}{C_2^2 \epsilon} x^\alpha$ , where  $\alpha = 2 + \deg p$ . The exact solution can be expressed using the Lambert  $W$  function [40]. Here, we only need the asymptotic expansion. Denote  $\beta = \frac{C_1}{C_2^2 \epsilon}$ , take logarithm, we have the



following iteration:

$$\begin{aligned} x &= \ln \beta + \alpha \ln x \\ &= \ln \beta + \alpha \ln(\ln \beta + \alpha \ln x) \\ &= \dots \\ &= O(\ln \beta). \end{aligned} \quad (\text{A5})$$

Thus, it is enough to choose  $r = O(\frac{1}{C_2} \ln \frac{C_1}{C_2^\alpha})$ , and  $N < 2d = O(r^2) = O(\frac{1}{C_2^2} \ln^2 \frac{C_1}{C_2^\alpha})$ . ■

As a corollary, it follows that

$$\begin{aligned} \sum_{|a| < \epsilon} |a| &= \sum_{|a| < \epsilon} \int_0^\epsilon \theta(|a| - x) dx = \int_0^\epsilon \sum_{|a| < \epsilon} \theta(|a| - x) dx \\ &< \int_0^\epsilon \frac{1}{C_2^2} \ln^2 \frac{C_1}{C_2^\alpha x} dx, \end{aligned} \quad (\text{A6})$$

which converges to 0 as  $\epsilon \rightarrow 0$ . Therefore, any EDP operator is trace class.

**Lemma 3.** Assume  $A$  is Hermitian. If  $\exists x \neq 0$  s.t.  $\|(A - \lambda)x\| < \|\epsilon x\|$ , then  $A$  has an eigenvalue  $a \in (\lambda - 2\epsilon, \lambda + 2\epsilon)$ . Moreover, decompose  $x = x^\parallel + x^\perp$  with respect to subspace  $(\lambda - 2\epsilon, \lambda + 2\epsilon)$ , then  $\|x^\perp\| < \frac{1}{2}$ . If  $A$  is of finite size  $N \times N$ , then an eigenvector  $x_a$  of  $A$  with eigenvalue  $a \in (\lambda - 2\epsilon, \lambda + 2\epsilon)$  satisfies  $|(x, y)| > \sqrt{\frac{3}{4N}}$ .

*Proof.* Denote  $B = A - \lambda$ , then  $\|Bx\| < \|\epsilon x\|$ . Without loss of generality, assume  $\|x\| = 1$ . Let us diagonalize  $B$ , so that  $B = \text{diag}\{b_1, \dots, b_n\}$ . Then, we have

$$\begin{aligned} \epsilon^2 &> \sum b_i^2 |x_i|^2 \\ &= \sum_{|b_i| \geq 2\epsilon} b_i^2 |x_i|^2 + \sum_{|b_i| < 2\epsilon} b_i^2 |x_i|^2 \\ &> 4\epsilon^2 \sum_{|b_i| \geq 2\epsilon} |x_i|^2 + \sum_{|b_i| < 2\epsilon} b_i^2 |x_i|^2. \end{aligned} \quad (\text{A7})$$

So, we must have  $\sum_{|b_i| \geq 2\epsilon} |x_i|^2 < \frac{1}{4}$ , thus  $\|x^\perp\| < \frac{1}{2}$  and  $\exists i$  such that  $|b_i| < 2\epsilon$ . If  $N$  is finite, then from  $\sum_{|b_i| < 2\epsilon} |x_i|^2 < \frac{3}{4}$  we know  $\exists i$  such that  $|b_i| < 2\epsilon$ ,  $|x_i| > \sqrt{\frac{3}{4N}}$ . ■

Go back to the original proposition. We want to find a correspondence between vertex eigenvalues of  $S$  and those of  $Q_1$  and  $Q_2$ . To make the notation simple, in the following  $e^{-r}$  means  $C_1 p(r) e^{-C_2 r}$  and  $C_1, C_2$  can change.

For each vertex state  $x$  of  $Q_1$ ,  $Q_1 x = qx$ ,  $\|x\| = 1$  ( $q \neq 0, 1$ ), separate  $x$  as  $x = y + z$  with respect to the disk  $B(1, r)$ . As in Eq. (A2), we still have  $\|z\| < \frac{e^{-r}}{|q^2 - q|}$ . It is not difficult to show that

$$\begin{aligned} \|(S - q)x\| &= \|(S - Q_1)x\| \\ &\leq \|(S - Q_1)y\| + \|(S - Q_1)z\| \\ &\lesssim \left(1 + \frac{1}{|q^2 - q|}\right) e^{-r} \stackrel{\text{def}}{=} \delta. \end{aligned} \quad (\text{A8})$$

According to Lemma 3,  $S$  has an eigenvalue in  $(q - 2\delta, q + 2\delta)$ . The same argument also applies to  $Q_2$ . Thus, for each vertex eigenvalue  $q$  of  $Q_1, Q_2$ , we get an eigenvalue of  $S$  in a neighborhood of  $q$ .

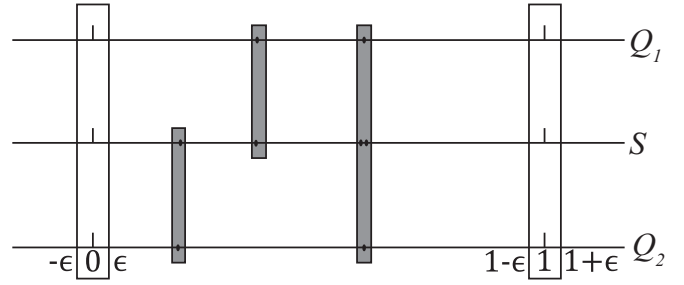


FIG. 7. Spectra structure of  $Q_i$  and  $S$ . We only consider eigenvalues outside  $(-\epsilon, \epsilon) \cup (1 - \epsilon, 1 + \epsilon)$ . The dots  $\bullet$  are eigenvalues. Each dot represents a Kramers pair due to time-reversal symmetry. The shaded windows are of width  $\delta \sim \frac{e^{-r}}{\epsilon}$  and are (from left to right) of three types.

Denote  $U_\epsilon \stackrel{\text{def}}{=} (-\epsilon, \epsilon) \cup (1 - \epsilon, 1 + \epsilon)$ . For  $q \notin U_\epsilon$ ,  $|q^2 - q| > \epsilon/2$ , so that  $\delta < \frac{e^{-r}}{\epsilon}$ . As  $r \rightarrow \infty$ , we will adjust  $\epsilon$  accordingly so that  $\delta$  is still small enough such that spectral gaps outside  $U_\epsilon$  are always greater than  $\delta$ . Then, we can describe the spectra structure of  $Q_1, Q_2$  in Fig. 7. The shaded windows have width  $\sim \delta$  and are of three types. For types 1 and 2, we already get the correspondence. For type 3, we claim the dimension of the subspace  $X$  corresponding to eigenvalues (of  $S$ ) in such window is at least 4. Indeed, denote  $x_1, x_2$  the eigenstates of  $Q_1$ ,  $x_3, x_4$  the eigenstates of  $Q_2$ , then  $|(x_i, x_j)| < \delta$ . Let  $x_i = u_i + v_i$  where  $u_i \in X$  and  $v_i \perp X$ , then,

$$|(y_i, y_j)| \leq |(x_i, x_j)| + |(z_i, z_j)| < \delta + \frac{1}{4}. \quad (\text{A9})$$

Similar to Eq. (A1) (here  $N = 4$ ), we get  $d \geq 4$ .

Moreover, each eigenstate of  $S$  with  $s \in U_\epsilon$  is generated in this way. Indeed, assume  $Sx = sx$ ,  $\|x\| = 1$  ( $s \neq 0, 1$ ), then,

$$x = \frac{1}{s^2 - s} (V_1 x + V_2 x) \quad (\text{A10})$$

is a decomposition of  $x$ . At least one of  $\|V_i x\|$  should be larger than  $|s^2 - s|/2$ , say  $V_1 x$ . Note that  $V_i x$  and  $(S - s)V_i x$  are mainly supported near vertex  $i$ , and

$$(S - s)V_1 x + (S - s)V_2 x = (S - s)(S^2 - S)x = 0, \quad (\text{A11})$$

and both terms must be small:

$$\|(S - s)V_i x\| < e^{-r}. \quad (\text{A12})$$

Therefore,

$$\begin{aligned} \|(Q_1 - s)V_1 x\| &\leq \|(S - s)V_i x\| + \|(S - Q_1)V_1 x\| \\ &< e^{-r} \\ &\leq \frac{e^{-r}}{|s^2 - s|} \|V_1 x\| \stackrel{\text{def}}{=} \delta' \|V_1 x\|. \end{aligned} \quad (\text{A13})$$

According to Lemma 3,  $Q_1$  has an eigenvalue in  $(s - 2\delta', s + 2\delta')$ . Therefore,  $s$  must be near (within  $\sim \delta$ ) a window, and an argument similar to (A9) shows that  $x$  cannot be a new eigenstate.



Due to this correspondence, the last term in the decomposition

$$\begin{aligned} & |\text{Tr}^v(S) - [\text{Tr}^v(Q_1) + \text{Tr}^v(Q_2)]| \\ & \leq \left| \sum_{s \in U_\epsilon} s \right| + \left| \sum_{q_1 \in U_\epsilon} q_1 \right| + \left| \sum_{q_2 \in U_\epsilon} q_2 \right| \\ & \quad + \left| \sum_{s \notin U_\epsilon} s - \sum_{q_1 \notin U_\epsilon} q_1 - \sum_{q_2 \notin U_\epsilon} q_2 \right|, \end{aligned} \quad (\text{A14})$$

is bounded by  $\delta \ln^2 \frac{1}{\epsilon}$  and goes to 0 as  $r \rightarrow \infty$ . The second and third terms can be bounded due to Eq. (A6) since  $Q_i - \bar{Q}_i$  obeys EDP with the same  $C_1$  and  $C_2$ . For the first term,  $S^2 - S$  also obeys EDP (with respect to the central point 0), however, with a new constant  $C'_1 \sim C_1 e^r$ . This is because [see Eq. (37)] the EDP of  $V_i$  is with respect to vertex  $i$ , so the decay property of  $S^2 - S$  with respect to vertex 0 needs to be estimated by  $e^{-\text{dist}(\mathbf{x}, 0)} < e^r e^{-\text{dist}(\mathbf{x}, 1)}$ . Luckily, similar to Eq. (A6), we still have

$$\left| \sum_{s \in U_\epsilon} s \right| < \int_0^\epsilon \frac{1}{C_2^2} \ln^2 \frac{C_1 e^r}{C_2^2 x} dx \rightarrow 0, \quad (\text{A15})$$

as  $r \rightarrow \infty$  as long as  $\epsilon = o(\frac{1}{r^2})$ . Thus, we have proved that

$$\lim_{r \rightarrow \infty} \text{Tr}^v(S) - [\text{Tr}^v(Q_1) + \text{Tr}^v(Q_2)] = 0. \quad (\text{A16})$$

Note that we have assumed that the requirement  $\epsilon = o(\frac{1}{r^2})$  is compatible with  $\delta = \frac{e^{-r}}{\epsilon} < \text{gap}$ . This technical assumption is reasonable. Indeed, according to Lemma 2, the spectral gap at  $\epsilon$  is roughly  $(\frac{d}{d\epsilon} \ln^2 \frac{1}{\epsilon})^{-1} \sim \frac{\epsilon}{\ln \epsilon}$  in average. In order for  $\delta < \frac{\epsilon}{\ln \epsilon}$ , it is enough to set  $\epsilon > \Omega(e^{-C'r})$ , which is exponentially smaller than  $\frac{1}{r^2}$  for large  $r$ . Even if we consider the fluctuation of the spectral gaps and even if the level statistics is Poissonian (so that no level repulsion), the probability for this to be true is 1 from the following estimation:

$$\begin{aligned} & \Pr \left( \text{at least one gap} < \frac{p(r)e^{-r}}{\epsilon} \right) \\ & < \sum_{x > \epsilon} \frac{p(r)e^{-r}/\epsilon}{x/\ln x} \\ & < \frac{p(r)e^{-r} \ln \epsilon}{\epsilon^2} \times \ln^2 \frac{1}{\epsilon} \rightarrow 0. \end{aligned} \quad (\text{A17})$$

## 2. Proof of the finite-size approximation

In this section, we prove Proposition 4. The technique used will be similar to the above section. We need to compare vertex eigenvalues of  $Q$  and  $Q'_N$ . Recall that  $\|Q'_N - Q_N\| \leq \rho^2$ ,  $\|V'_N - W_N\| < \rho$ ,  $\|V'_N - \bar{W}_N\| < \rho$  where  $\rho \sim e^{-Cr}$ .

Temporarily fix  $\epsilon$ , and only consider eigenvalues outside  $U_\epsilon \stackrel{\text{def}}{=} (-\epsilon, \epsilon) \cup (1 - \epsilon, 1 + \epsilon)$ . For any  $\alpha \notin U_\epsilon$ ,  $|\alpha^2 - \alpha| > \epsilon/2$ .

For  $(Q - q)x = 0$ ,  $q \notin U_\epsilon$ , we separate it as  $x = y + z$  with respect to circle  $r/2$ , again  $\|z\| < \frac{p(r)e^{-Cr}}{\epsilon} \stackrel{\text{def}}{=} \delta$ . In the following,  $\delta$  means “everything that goes like  $\frac{p(r)e^{-Cr}}{\epsilon}$ ” with

perhaps different  $C$ .” Thus,

$$\begin{aligned} \| (Q'_N - q)x \| &= \| (Q'_N - Q_N + Q_N - Q)x \| \\ &\leq \| (Q'_N - Q_N)x \| + \| (Q_N - Q)y \| \\ &\quad + \| (Q_N - Q)z \| \lesssim \delta, \end{aligned} \quad (\text{A18})$$

so  $Q'_N$  has an eigenvalue  $q' \in (q - 2\delta, q + 2\delta)$  with eigenstate  $x'$  satisfies  $|(x, x')| > \sqrt{\frac{3}{4N}}$  (Lemma 3). This implies  $x'$  must contain a vertex eigenstate. Indeed, if not, we have  $V'_N x' = 0$  so  $x' = \frac{1}{q^2 - q'} W'_N x'$  which is concentrated near boundary  $r$ , thus,

$$\begin{aligned} |(x, x')| &= \frac{1}{|q'^2 - q'|} |(x, W'_N x')| \\ &< \frac{2}{\epsilon} [| (y, W'_N x') | + | (z, W'_N x') |] \lesssim \delta, \end{aligned} \quad (\text{A19})$$

a contradiction as  $r \rightarrow \infty$ .

On the other hand, if  $(Q'_N - a)x = 0$  ( $a \neq 0, 1$ ), and  $x$  is a vertex state:  $W'_N x = 0$ , then  $x = \frac{V'_N x}{a^2 - a}$ . We have

$$\begin{aligned} (Q - a)x &= \frac{1}{a^2 - a} (Q - Q'_N) V'_N x \\ &= \frac{1}{a^2 - a} [(Q - Q_N) V_N x + (Q_N - Q'_N) V'_N x] \lesssim \delta. \end{aligned} \quad (\text{A20})$$

So,  $Q$  has an eigenvalue in  $(a - 2\delta, a + 2\delta)$ .

Now, we choose  $r$  according to the same technical assumption above, so that there is a correspondence outside region  $U_\epsilon$  for  $\forall \epsilon$ . Then, similarly we have

$$\left| \sum_{\text{vertex}} q' - \text{Tr}^v(Q) \right| \leq \left| \sum_{q \in U_\epsilon} q \right| + \left| \sum_{q' \in U_\epsilon} q' \right| + \left| \sum_{q \notin U_\epsilon} q - \sum_{q' \notin U_\epsilon} q' \right|. \quad (\text{A21})$$

The last term is bounded by  $\delta \ln^2 \frac{1}{\epsilon}$  which goes to 0 as  $\epsilon \rightarrow 0$ . The first term also goes to 0 since  $Q - \bar{Q}$  is trace class. The second term is bounded by  $\#\{q'\} \epsilon$ . Obviously,  $\#\{q'\} < \dim Q_N \sim r^2$ , so this term also converges to 0 since  $\epsilon = o(\frac{1}{r^2})$ .

## 3. Proof that $S - T$ is trace class

In this section, we prove the claim used in Property 4. The first step is to figure out the decay behavior of the matrix elements of  $S - T$ . According to the Peierls substitution [36],

$$S_{ij} = P_{ij} e^{i \int_i^j \mathbf{A} \cdot d\mathbf{r}}, \quad T_{ij} = P_{ij} e^{i \int_i^j \mathbf{A}' \cdot d\mathbf{r}}, \quad (\text{A22})$$

where  $\mathbf{A}$  and  $\mathbf{A}'$  are the vector potentials for the two flux configurations [see Fig. 8(a) (all angles here are directed)]; we have

$$\begin{aligned} \mathbf{A} \cdot d\mathbf{r} &\propto 2\theta, \quad \mathbf{A}' \cdot d\mathbf{r} \propto \theta_1 + \theta_2, \\ (S - T)_{\mathbf{x}, \mathbf{y}} &\sim P_{\mathbf{x}, \mathbf{y}} e^{i(\theta_1 + \theta_2)} (e^{i(2\theta - \theta_1 - \theta_2)} - 1). \end{aligned} \quad (\text{A23})$$



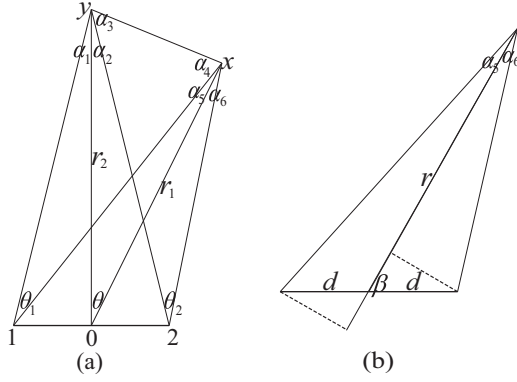


FIG. 8. Relevant geometries in the proof.

From geometry,  $2\theta - \theta_1 - \theta_2 = (\alpha_1 - \alpha_2) - (\alpha_5 - \alpha_6)$ . Let us calculate  $(\alpha_5 - \alpha_6)$  [see Fig. 8(b)]; we have

$$\begin{aligned} \alpha_5 - \alpha_6 &= \arctan \frac{d \sin \beta}{r + d \cos \beta} - \arctan \frac{d \sin \beta}{r - d \cos \beta} \\ &= \arctan \frac{\frac{d \sin \beta}{r + d \cos \beta} - \frac{d \sin \beta}{r - d \cos \beta}}{1 + \frac{d \sin \beta}{r + d \cos \beta} \frac{d \sin \beta}{r - d \cos \beta}} \\ &= -\arctan \frac{d^2 \sin 2\beta}{r^2 - d^2} \\ &= -\frac{d^2 \sin 2\beta}{r^2} + O\left(\frac{1}{r^4}\right). \end{aligned} \quad (\text{A24})$$

Due to the energy gap,  $|A_{\mathbf{x},\mathbf{y}}| \lesssim e^{-C|\mathbf{x}-\mathbf{y}|}$ . Assuming  $r_1 \geq r_2$ , we claim that

$$e^{-C|\mathbf{x}-\mathbf{y}|} |e^{i(2\theta - \theta_1 - \theta_2)} - 1| < e^{-\frac{C}{2}|\mathbf{x}-\mathbf{y}|} O\left(\frac{1}{r_1^3}\right). \quad (\text{A25})$$

Indeed, if  $e^{-\frac{C}{2}|\mathbf{x}-\mathbf{y}|} < \frac{1}{(r_1+r_2)^3}$ , there is nothing needed to prove. If not, then  $|\mathbf{x}-\mathbf{y}| < \frac{6}{C} \ln(r_1+r_2) < \frac{7}{C} \ln r_1$  (asymptotically). In this case, from geometry, we know  $|\beta_i - \beta_j| = |\theta| \lesssim \frac{|\mathbf{x}-\mathbf{y}|}{r_1}$ . Then,  $2\theta - \theta_1 - \theta_2 = (\alpha_1 - \alpha_2) - (\alpha_5 - \alpha_6) = \frac{\sin 2\beta_1}{r_1^2} - \frac{\sin 2\beta_2}{r_2^2} + O\left(\frac{1}{r_1^4}\right)$  will be of order  $O\left(\frac{|\mathbf{x}-\mathbf{y}|}{r_1^3}\right)$  as can be seen from Taylor expansion. Then, it is easy to see that the claim also holds.

The result is (in a more symmetric fashion, ignore constants) as follows: the operator  $A = S - T$  satisfies the decay property

$$|A_{\mathbf{x},\mathbf{y}}| < \frac{1}{(|\mathbf{x}| + |\mathbf{y}|)^3} e^{-|\mathbf{x}-\mathbf{y}|}. \quad (\text{A26})$$

Now, we prove this kind of operator must be trace class. Let us denote the  $n$ th singular value (decreasing order) to be  $s_n$ . According to the Courant min-max principle [34], we have

$$s_n = \min_{Y_{n-1}} \max_{u \perp Y_{n-1}} \frac{(Au, Au)}{(u, u)}, \quad (\text{A27})$$

where  $Y_{n-1}$  means a subspace of dimension  $n-1$ . Thus, for any given  $n$ -dimensional subspace  $Y_{n-1}$ , we have

$$s_n^2 \leq \max_{u \perp Y_{n-1}} \frac{(Au, Au)}{(u, u)} = \max_{\substack{u \perp Y_{n-1} \\ \|u\|=1}} \|Au\|^2. \quad (\text{A28})$$

Let us choose the subspace  $Y_{n-1}$  to be spanned by the  $n$  components nearest to the center (so that the label of the components are approximately in the disk of radius  $r \sim \sqrt{n}$ ). Denote the columns of  $A$  to be  $v_{\mathbf{x}}$  ( $v_{\mathbf{x}} = Ae_{\mathbf{x}}$ ,  $\mathbf{x} \in \mathbb{Z}^2$  is the label). With Eq. (A26) it is easy to show (note that here  $e^{-|\mathbf{x}-\mathbf{y}|}$  means  $e^{-C|\mathbf{x}-\mathbf{y}|}$  for a different  $C$ )

$$|(v_{\mathbf{x}}, v_{\mathbf{y}})| = |(A^2)_{\mathbf{x},\mathbf{y}}| \lesssim \frac{e^{-|\mathbf{x}-\mathbf{y}|}}{|\mathbf{x}|^3 |\mathbf{y}|^3}. \quad (\text{A29})$$

Thus,

$$\begin{aligned} \|Au\|^2 &= \left\| \sum_{|\mathbf{x}| > r} u_{\mathbf{x}} v_{\mathbf{x}} \right\|^2 \\ &= \sum_{|\mathbf{x}|, |\mathbf{y}| > r} \bar{u}_{\mathbf{x}} u_{\mathbf{y}} (v_{\mathbf{x}}, v_{\mathbf{y}}) \\ &= \left( \sum_{\substack{|\mathbf{x}-\mathbf{y}| \geq l \\ |\mathbf{x}|, |\mathbf{y}| > r}} + \sum_{\substack{|\mathbf{x}-\mathbf{y}| < l \\ |\mathbf{x}|, |\mathbf{y}| > r}} \right) \bar{u}_{\mathbf{x}} u_{\mathbf{y}} (v_{\mathbf{x}}, v_{\mathbf{y}}). \end{aligned} \quad (\text{A30})$$

Here,  $l$  will be of the order  $\ln r$ , to be specific later.

The first summation is (crude but enough) controlled by  $e^{-l}$  due to Eq. (A29) and Cauchy inequality. For the second summation, we have

$$\left| \sum \bar{u}_{\mathbf{x}} u_{\mathbf{y}} (v_{\mathbf{x}}, v_{\mathbf{y}}) \right| < \frac{1}{4} \sum (|u_{\mathbf{x}}|^2 + |u_{\mathbf{y}}|^2) |(v_{\mathbf{x}}, v_{\mathbf{y}})| \lesssim \frac{l^2}{r^6}. \quad (\text{A31})$$

Let us choose  $l$  such that  $e^{-l} = \frac{1}{r^6}$ , and we finally have

$$s_n^2 \leq e^{-l} + \frac{l^2}{r^6} < \frac{\ln^2 r}{r^6} \sim \frac{\ln^2 n}{n^3}, \quad (\text{A32})$$

so  $\sum s_n = \sum \frac{\ln n}{n^{3/2}}$  converges, which means  $A$  is trace class.

- [1] A. Altland and M. R. Zirnbauer, *Phys. Rev. B* **55**, 1142 (1997).
- [2] A. P. Schnyder, S. Ryu, A. Furusaki, and A. W. W. Ludwig, *Phys. Rev. B* **78**, 195125 (2008).
- [3] A. Kitaev, in *American Institute of Physics Conference Proceedings Series*, American Institute of Physics Conference Series, Vol. 1134, edited by V. Lebedev and M. Feigel'Man (AIP, Melville, NY, 2009), pp. 22–30.
- [4] D. S. Freed and G. W. Moore, *Ann. Henri Poincaré* **14**, 1927 (2013).

- [5] R. Bott, *Proc. Natl. Acad. Sci. USA* **43**, 933 (1957).
- [6] R. Bott, *Ann. Math.* **70**, 313 (1959).
- [7] C. L. Kane and E. J. Mele, *Phys. Rev. Lett.* **95**, 146802 (2005).
- [8] C. L. Kane and E. J. Mele, *Phys. Rev. Lett.* **95**, 226801 (2005).
- [9] L. Fu and C. L. Kane, *Phys. Rev. B* **74**, 195312 (2006).
- [10] L. Fu and C. L. Kane, *Phys. Rev. B* **76**, 045302 (2007).
- [11] J. E. Moore and L. Balents, *Phys. Rev. B* **75**, 121306(R) (2007).
- [12] X.-L. Qi, T. L. Hughes, and S.-C. Zhang, *Phys. Rev. B* **78**, 195424 (2008).



- [13] R. Roy, *Phys. Rev. B* **79**, 195321 (2009).
- [14] J. Bellissard, in *Statistical Mechanics and Field Theory: Mathematical Aspects*, Lecture Notes in Physics, Vol. 257, edited by T. C. Dorlas, N. M. Hugenholtz, and M. Winnink (Springer, Berlin, 1986), pp. 99–156.
- [15] J. Bellissard, A. van Elst, and H. Schulz-Baldes, *J. Math. Phys.* **35**, 5373 (1994).
- [16] A. Connes, *Noncommutative Geometry* (Elsevier, Amsterdam, 1995).
- [17] E. Prodan, *New J. Phys.* **12**, 065003 (2010).
- [18] E. Prodan, *J. Phys. A: Math. Gen.* **44**, 113001 (2011).
- [19] E. Prodan, B. Leung, and J. Bellissard, *J. Phys. A* **46**, 485202 (2013).
- [20] J. E. Avron, R. Seiler, and B. Simon, *Commun. Math. Phys.* **159**, 399 (1994).
- [21] T. A. Loring and M. B. Hastings, *Europhys. Lett.* **92**, 67004 (2010).
- [22] M. B. Hastings and T. A. Loring, *J. Math. Phys.* **51**, 015214 (2010).
- [23] M. B. Hastings and T. A. Loring, *Ann. Phys.* **326**, 1699 (2011), July 2011 Special Issue.
- [24] T. A. Loring, *Ann. Phys.* **356**, 383 (2015).
- [25] G. De Nittis and H. Schulz-Baldes, *Ann. Henri Poincaré* **17**, 1 (2016).
- [26] H. Katsura and T. Koma, *J. Math. Phys.* **57**, 021903 (2016).
- [27] Y. Akagi, H. Katsura, and T. Koma, *J. Phys. Soc. Jpn.* **86**, 123710 (2017).
- [28] R. Bianco and R. Resta, *Phys. Rev. B* **84**, 241106(R) (2011).
- [29] H. Huang and F. Liu, *Phys. Rev. Lett.* **121**, 126401 (2018).
- [30] I. C. Fulga, F. Hassler, and A. R. Akhmerov, *Phys. Rev. B* **85**, 165409 (2012).
- [31] A. Kitaev, *Ann. Phys.* **321**, 2 (2006), January Special Issue.
- [32] N. P. Mitchell, L. M. Nash, D. Hexner, A. M. Turner, and W. T. M. Irvine, *Nat. Phys.* **14**, 380 (2018).
- [33] D. J. Thouless, *Phys. Rev. B* **27**, 6083 (1983).
- [34] P. Lax, *Functional Analysis*, Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts (Wiley, Hoboken, NJ, 2014).
- [35] M. B. Hastings and T. Koma, *Commun. Math. Phys.* **265**, 781 (2006).
- [36] R. Peierls, *Z. Phys.* **80**, 763 (1933).
- [37] H. Lin, *Fields Institute Commun.* **13**, 193 (1997).
- [38] M. B. Hastings, *Commun. Math. Phys.* **291**, 321 (2009).
- [39] B. A. Bernevig, T. L. Hughes, and S.-C. Zhang, *Science* **314**, 1757 (2006).
- [40] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, *Adv. Comput. Math.* **5**, 329 (1996).

*Correction:* A few phrases in text were modified incorrectly during production and have been restored.